Integrazioni del corso di Calcolo delle Probabilità
Università degli Studi di Roma La Sapienza
Docente: Alessandra Faggionato

Contents

1. Spazio ingenuo di probabilità

Dato un insieme S, denotiamo con $\mathcal{P}(S)$ l'insieme delle parti di S.

Definizione 1. *Uno spazio (ingenuo) di probabilità è una coppia $(S,\mathbb{P})$, dove S è un insieme non vuoto e $\mathbb{P}$ un'applicazione $\mathbb{P}\colon \mathcal{P}(S)\to[0,1]$ tale che*

(a) $\mathbb{P}(S)=1$.
(b) *Per ogni famiglia numerabile $\{E_i\}_{i=1}^\infty$ di sottoinsiemi di S a due a due disgiunti (ossia $E_i\cap E_j=\emptyset$ per $i\neq j$) si ha*

$$\mathbb{P}\Big(\bigcup_{i=1}^\infty E_i\Big)=\sum_{i=1}^\infty \mathbb{P}(E_i)$$

In questo contesto, S si dice uno spazio campionario*, i sottoinsiemi di S* eventi *e $\mathbb{P}$ una* misura *o* legge di probabilità *su S.*

Remark 1. *Purtroppo tale definizione è troppo restrittiva, e molti modelli probabilistici interessanti non rientrano nella definizione di spazio (ingenuo) di probabilità data sopra. In generale è sufficiente che $\mathbb{P}$ sia una funzione $\mathbb{P}\colon\mathfrak{F}\to[0,1]$, dove $\mathfrak{F}$ è un sottoinsieme di $\mathcal{P}(S)$ con opportune proprietà (σ-algebra).*

Proposizione 1. *Se S è numerabile (finito o infinito) vale*

$$P(E)=\sum_{a\in E}P(\{a\})\qquad \forall E\subset S\,. \tag{1.1}$$

Proof. Possiamo scrivere E come l'unione disgiunta $E=\cup_{a\in E}\{a\}$ di una famiglia numerabile di eventi elementari. Invocando l'additività finita se E è finito, e l'additività numerabile se E è infinito numerabile concludiamo che vale (1.1). □

Proposizione 2. *Data una famiglia finita di eventi $E_1, E_2, \dots, E_n$, vale $\mathbb{P}(\cup_{i=1}^n E_i) \leq \sum_{i=1}^n \mathbb{P}(E_i)$. Inoltre le stessa proprietà vale per una successione di eventi.*

Proof. Trattiamo solo il caso della successione di eventi, il caso di famiglia finita e' completamente analogo. Definiamo

$$F_1 := E_1\,,\ F_i = E_i \setminus \cup_{k=1}^{i-1} E_k \quad \text{per } i \geq 2\,.$$

Per costruzione $F_1, F_2, F_3, \dots$ è una successione di eventi a due a due disgiunti. Per l'additività numerabile abbiamo

$$P\left(\cup_{i=1}^\infty F_i\right) = \sum_{i=1}^\infty P(F_i)\,. \tag{1.2}$$

Osservando che $\cup_{i=1}^\infty F_i = \cup_{i=1}^\infty E_i$ e che per la proprietà di monotonia $P(F_i) \leq P(E_i)$ essendo $F_i \subset E_i$, da (1.2) otteniamo che $P\left(\cup_{i=1}^\infty E_i\right) \leq \sum_{i=1}^\infty P(E_i)$. □

Proposizione 3 (Principio di inclusione esclusione)**.** *Dati $E_1, \dots, E_n$ eventi, vale*

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{r=1}^n (-1)^{r+1} \sum_{\substack{(i_1,i_2,\dots,i_r):\\ 1 \leq i_1 < i_2 < \cdots < i_r \leq n}} P(E_{i_1} E_{i_2} \cdots E_{i_r})\,. \tag{1.3}$$

Proof. Diamo la dimostrazione con S numerabile (chi non ha familiarità con le serie assolutamente convergenti può limitarsi a S finito). Notiamo che (1.1) può essere riscritta come

$$P(E) := \sum_{a \in S} P(\{a\}) \mathbb{1}(a \in E)$$

con $\mathbb{1}$ funzione caratteristica. Quindi scrivendo

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{a \in S} P(\{a\}) \mathbb{1}(a \in E_1 \cup E_2 \cup \cdots \cup E_n)\,, \tag{1.4}$$

$$P(E_{i_1} E_{i_2} \cdots E_{i_r}) = \sum_{a \in S} P(\{a\}) \mathbb{1}(a \in E_{i_1} E_{i_2} \cdots E_{i_r})\,, \tag{1.5}$$

(1.3) puo' essere riscritta come

$$\sum_{a \in S} P(\{a\}) \mathbb{1}(a \in E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{a \in S} P(\{a\}) \Big[\sum_{r=1}^n (-1)^{r+1} \sum_{\substack{(i_1,i_2,\dots,i_r):\\ 1 \leq i_1 < i_2 < \cdots < i_r \leq n}} \mathbb{1}(a \in E_{i_1} E_{i_2} \cdots E_{i_r}) \Big]\,. \tag{1.6}$$

Per verificare (1.6) ci basta provare per ogni $a \in S$ che

$$\mathbb{1}(a \in E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{r=1}^n (-1)^{r+1} \sum_{\substack{(i_1,i_2,\dots,i_r):\\ 1 \leq i_1 < i_2 < \cdots < i_r \leq n}} \mathbb{1}(a \in E_{i_1} E_{i_2} \cdots E_{i_r}) \tag{1.7}$$

Se $a \notin E_1 \cup E_2 \cup \cdots \cup E_n$ la suddetta identità è banalmente vera dato che si riduce a $0 = 0$. Sia $a \in E_1 \cup E_2 \cup \cdots \cup E_n$. Supponiamo che a appartenga ad esattamente m insiemi del tipo E_i. Allora

$$\sum_{\substack{(i_1,i_2,\dots,i_r):\\ 1 \leq i_1 < i_2 < \cdots < i_r \leq n}} \mathbb{1}(a \in E_{i_1} E_{i_2} \cdots E_{i_r})$$

corrisponde a quanti modi possiamo scegliere r indici tra gli m indici i per cui $a \in E_i$, e quindi vale $\binom{m}{r}$. Ne deriva che (1.7) equivale a

$$1 = \sum_{r=1}^{m} (-1)^{r+1} \binom{m}{r}$$

ovvero (portando tutto a sx)

$$\sum_{r=0}^{m} (-1)^{r} \binom{m}{r} = 0.$$

La suddetta indetità è vera dato che il lato sx è lo sviluppo del binomio $(-1+1)^m$. □

Proposizione 4. *Sia S insieme numerabile (finito o infinito). Allora valgono i seguenti fatti:*

(i) *se P è una funzione di probabilità avente spazio campionario S, allora la funzione $p : S \to \mathbb{R}$ definita come $p(s) := P(\{s\})$ per $s \in S$ soddisfa*

$$p(s) \geq 0 \qquad \forall s \in S\,, \qquad \sum_{s\in S} p(s) = 1\,. \tag{1.8}$$

(ii) *Viceversa, data una funzione $p : S \to \mathbb{R}$ che soddisfa* (1.8)*, esiste un'unica funzione di probabilità P avente S spazio campionario per cui $P(\{s\}) = p(s)$. Inoltre, vale*

$$P(E) := \sum_{s\in E} p(s)\,, \qquad \forall E \subset S\,. \tag{1.9}$$

Proof. Ci limitiamo al caso di S finito. La dimostrazione nel caso generale è simile, ed usa alcune proprietà delle serie a segno nonnegativo.

Proviamo il punto (i). Sia $p(s) = P(\{s\})$. Tale numero per assioma è in $[0,1]$. Siccome S è finito, possiamo scrivere $S = \{s_1, s_2, \dots, s_n\}$. Quindi $S = \{s_1\} \cup \{s_2\} \cup \dots \cup \{s_n\}$. Siccome gli eventi elementari $\{s_1\}$, $\{s_2\}$,... $\{s_n\}$ sono a due a due disgiunti, per l'additività finita vale

$$P(S) = \sum_{i=1}^{n} P(\{s_i\})\,.$$

Notiamo che il membro sinistro è 1 per assioma, il membro destro si puo' riscrivere come $\sum_{i=1}^{n} p(s_i)$, quindi abbiamo completato la dimostrazione di (1.8).

Proviamo il punto (ii). Sia P una funzione di probabilità per cui $P(\{s\}) = p(s)$. Dato $E \subset S$, scrivendo E come l'unione finita e disgiunta degli eventi elementari $\{s\}$ con $s \in E$, ovvero

$$E = \cup_{s\in E}\{s\}\,,$$

per l'additività finita abbiamo $P(E) = \sum_{s\in E} P(\{s\}) = \sum_{s\in E} p(s)$. Quindi necessariamente P verifica (1.9). Verifichiamo ora che la funzione P definita sull'insieme delle parti $\mathcal{P}(S)$ tramite (1.9) soddisfa gli assiomi. Poiche' gli addendi nel membro destro in (1.9) sono nonnegativi abbiamo che $P(E) \geq 0$ ed inoltre che $P(E) \leq \sum_{i=1}^{n} p(s_i) = 1$. Quindi abbiamo provato che $P(E) \in [0,1]$. Banalmente $P(S) = 1$ dato che $\sum_{i=1}^{n} p(s_i) = 1$. La numerabile additivià per S finito equivale all'additività finita e quest'ultima deriva dalla proprietà associativa della somma.

□

2. Esempi di spazi di probabilità con spazi campionari non numerabili

Esempio 1. *Modellizziamo l'esperimento ideale in cui si sceglie a caso un punto dall'intervallo* $[0,1]$*. Lo spazio campionario è allora* $S=[0,1]$*. Risulta naturale proporre come funzione di probabilità*

$$P(E) := \text{lunghezza di } E \qquad E \subset S$$

Siccome non tutti i sottinsiemi di $S=[0,1]$ *hanno una lunghezza ben definita, la funzione* P *non potrà essere definita per ogni* $E \subset [0,1]$*. Si prova infatti che non esiste una funzione di probabilità sull'insieme delle parti di* $[0,1]$ *per cui* $P(E)=b-a$ *se* $E \subset [0,1]$ *è l'intervallo* $[a,b]$*. Questo problema si supera modificando la definizione di spazio di probabilità per cui gli eventi non sono tutti i possibili sottinsiemi di* S*, ma una sottofamiglia con opportune proprietà (detta* σ*–algebra) per cui la lunghezza è ben definita. Si noti che per definizione* $P(\{s\})=0$ *per ogni esito* s*, cioè gli esiti hanno probabilità di realizzazione nulla (anche se uno verrà realizzato).*

Esempio 2. *Lanciamo una moneta onesta infinite volte, e ad ogni lancio identifichiamo "testa" con 0 e "croce" con 1. Lo spazio campionario è dato da* $S=\{0,1\}^{\mathbb{N}^+}=\{(x_1,x_2,\dots):x_k\in\{0,1\}\ \forall k=1,2,3,\dots\}$*.* x_k *è da intendersi come il risultato del lancio k-esimo. Dato* $a \in S$ *e* $n \in \mathbb{N}^+$*, consideriamo l'evento* E_a^n

$$E_a^n := \{x \in S \,:\, x_i = a_i,\ \forall i = 1,\dots,n\}$$

Se ad esempio $a=(0,0,0,\dots)$*, allora* E_a^n *è l'evento che corrisponde ad ottenere n volte "testa" nei primi n lanci.*

Anche se non sappiamo come definire una misura di probabilità su ogni sottoinsieme di S *(*S *ha la cardinalità del continuo), la misura di probabilità* $\mathbb{P}$ *corrispondente a lanci di una moneta "onesta" è tale che*

$$\mathbb{P}(E_a^n) = 2^{-n} \tag{2.1}$$

per ogni $a \in S$*. Si noti che* $E_a^n \subset E_a^{n'}$ *se* $n \geq n'$ *e che* $\{a\} = \cap_{n\geq 1} E_a^n$*. Pertanto, applicando il teorema del limite di successioni monotone abbiamo*

$$\mathbb{P}(\{a\}) = \mathbb{P}(\cap_{n\geq 1} E_a^n) = \lim_{n\to+\infty} \mathbb{P}(E_a^n) = \lim_{n\to+\infty} 2^{-n} = 0$$

Vale a dire, ogni punto di S *ha probabilità 0, e chiaramente* $\mathbb{P}$ *non è identificata dai suoi valori sui singleton.*

Esempio 3. *Dato un insieme non vuoto* S*, e fissato un elemento* $x \in S$*, consideriamo la misura di probabilità su* S *detta* delta di Dirac *(ed indicata con* $\mathbb{P}=\delta_x$*), definita da*

$$\delta_x(E) = \begin{cases} 1 & \text{se } x \in E \\ 0 & \text{se } x \notin E \end{cases} \qquad E \subset S$$

È immediato verificare che δ_x *è una misura di probabilità su* S*. Si può anche verificare che* δ_x *è l'unica misura di probabilità* $\mathbb{P}$ *su* S *tale che, per ogni* $y \in S$

$$\mathbb{P}(\{y\}) = \begin{cases} 1 & \text{se } x = y \\ 0 & \text{se } x \neq y \end{cases}$$

Ossia, anche nel caso di S *non numerabile,* δ_x *è identificata dai suoi valori sui singleton.*

3. Spazi di probabilità prodotto

Iniziamo la nostra discussione con un esempio: vogliamo modelizzare con uno spazio di probabilità l'esperimento che consiste nel lanciare prima un dado e poi una moneta truccata, dove la probabilità di avere testa è 1/3. I due lanci sono considerati operativamente indipendenti, cioè l'uno non puo' influenzare il risultato dell'altro. Lo spazio campionario S, dato dall'insieme dei possibili esiti, è banalmente

$$S = \left\{ \, (x_1, x_2) \, : x_1 \in \{1, 2, \ldots, 6\}, \, x_2 \in \{T, C\} \, \right\}$$

(x_1 è da pensare come l'esito del lancio del dado e x_2 come l'esito del lancio della moneta). Essendo S finito (quindi numerabile) ci resta solo da definire la funzione di probabilità $\mathbb{P} : \mathcal{P}(S) \to [0, 1]$. Siccome la moneta è truccata, non possiamo invocare la simmetria e concludere che gli esiti di S sono equiprobabili. Invece, essendo gli esperimenti operativamente indipendenti, è naturale richiedere che, fissati $x_1 \in \{1, 2, \ldots, 6\}$ e $x_2 \in \{T, C\}$, i due eventi (visti come sottinsiemi di S) E_1 e E_2, definiti come E_1="il lancio del dado ha dato x_1" e E_2="il lancio della moneta ha dato x_2", siano eventi indipendenti. Quindi P deve essere tale che

$$\mathbb{P}\big(\{(x_1, x_2)\}\big) = P(E_1 \cap E_2) = P(E_1) P(E_2) \, .$$

È naturale porre $\mathbb{P}(E_1) = 1/6$ e $\mathbb{P}(E_2) = 1/3$ se $x_2 = T$ oppure $\mathbb{P}(E_2) = 2/3$ se $x_2 = C$. Quindi possiamo definire $\mathbb{P}$ sugli eventi elementari come

$$\mathbb{P}\big(\{(x_1, x_2)\}\big) = \begin{cases} \frac{1}{6}\frac{1}{3} = \frac{1}{18} & \text{se } x_2 = T, \\ \frac{1}{6}\frac{2}{3} = \frac{2}{18} & \text{se } x_2 = C, \end{cases}$$

Ricordiamo che, essendo S numerabile, la funzione di probabilità $\mathbb{P}$ è univocalmente determinata una volta assegnata sugli eventi elementari. Per avere una buona definizione bisogna però verificare che $\mathbb{P}(\{x\}) \in [0, 1]$ per ogni $x \in S$ e che $\sum_{x \in S} \mathbb{P}(\{x\}) = 1$. Queste verifiche seguono facilmente dalla definizione e sono lasciate al lettore.

Il precedente esempio–guida puo' essere facilmente generalizzato come segue. Consideriamo un esperimento dato da n sotto–esperimenti tra loro indipendenti (ci riferiamo all'indipendenza operativa nell'esecuzione degli esperimenti elementari). Supponiamo che l'esperimento i–esimo sia descritto dallo spazio di probabilità $(S_i, \mathbb{P}_i)$, dove S_i è lo spazio campionario che assumiamo essere numerabile (finito o infinito), e $P_i : \mathcal{P}(S_i) \to [0, 1]$ è la funzione di probabilità.

Allora l'esperimento globale dato dagli n sotto–esperimenti elementari tra loro indipendenti è descritto dallo **spazio di probabilità** $(S, \mathbb{P})$ così definito:

- Lo spazio campionario S è dato dal prodotto cartesiano
$$S = S_1 \times S_2 \times \cdots \times S_n = \left\{ \, (x_1, x_2, \ldots, x_n) \, : \, x_i \in S_i \;\; \forall 1 \leq i \leq n \, \right\} .$$
- Essendo S numerabile, definiamo $\mathbb{P} : \mathcal{P}(S) \to [0, 1]$ come l'unica funzione di probabilità tale che
$$\mathbb{P}\left(\big\{(x_1, x_2, \ldots, x_n)\big\}\right) = \prod_{i=1}^{n} P_i(\{x_i\}) \, . \tag{3.1}$$

Essendo S numerabile la funzione di probabilità è univocalmente determinata una volta assegnata sugli eventi elementari. Per avere una buona definizione bisogna però verificare

che $\mathbb{P}(\{x\}) \geq 0$ per ogni $x \in S$ e che $\sum_{x\in S} \mathbb{P}(\{x\}) = 1$. Queste verifiche seguono facilmente dalla definizione (3.1) e sono lasciate al lettore.
La formula (3.1) è motivata dalla seguente osservazione: se E_i è l'evento che il sotto–esperimento i-esimo abbia esito x_i allora, essendo i sotto–esperimenti operativamente indipendenti, è naturale supporre che $E_1, E_2, \ldots, E_n$ siano eventi indipendenti. Inoltre è naturale assumere che $\mathbb{P}(E_i) = \mathbb{P}_i(\{x_i\})$. Quindi

$$\mathbb{P}\left(\left\{(x_1, x_2, \ldots, x_n)\right\}\right) = \mathbb{P}\left(\cap_{i=1}^n E_i\right) = \prod_{i=1}^n \mathbb{P}(E_i) = \prod_{i=1}^n \mathbb{P}_i(\{x_i\})\,.$$

Il suddetto spazio di probabilità $(S, \mathbb{P})$ è detto spazio di probabilità prodotto degli spazi di probabilità $(S_i, \mathbb{P}_i)$, $1 \leq i \leq n$.

Valgono le seguenti proprietà:

Proposizione 5. *Siano $(S_i, \mathbb{P}_i)$, $1 \leq i \leq n$, spazi di probabilità con S_i insieme numerabile, e sia $(S, \mathbb{P})$ il loro spazio prodotto.*
1) Se per ogni i $E_i \subset S$ è un evento definito in termini del solo sotto–esperimento i–esimo, allora $E_1, E_2, \ldots, E_n$ sono eventi indipendenti,
2) Se per ogni i S_i è finito ed ha esiti equiprobabiliti, allora S ha esiti equiprobabili.

Osserviamo che la prima delle suddette proprietà conferma che lo spazio prodotto è il giusto spazio di probabilità per modellizzare gli n esperimenti elementari indipendenti. Notiamo inoltre che se $E_i \subset S$ è un evento definito in termini del solo esperimento elementare i–esimo allora E_i deve essere della forma

$$E_i = S_1 \times S_2 \times S_{i-1} \times \tilde{E}_i \times S_{i+1} \cdots \times S_n, \tag{3.2}$$

per qualche insieme $\tilde{E}_i \subset S_i$.

Dim. della Proposizione 5.
Dimostriamo il primo punto. Per provare che $E_1, E_2, \ldots, E_n$ sono indipendenti dobbiamo provare che, dati $1 \leq i_1 < i_2 < \cdots < i_r \leq n$, vale

$$\mathbb{P}\left(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_r}\right) = \mathbb{P}(E_{i_1}) \cdot \mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_r})\,.$$

Per semplicità di notazione (ma il ragionamento si generalizza in modo banale) consideriamo solo il caso in cui $r = n$ e quindi $i_1 = 1, i_2 = 2, \ldots i_n = n$, cioè proviamo che

$$\mathbb{P}\left(\cap_{i=1}^n E_i\right) = \prod_{i=1}^n \mathbb{P}(E_i)\,. \tag{3.3}$$

Come appena osservato, E_i è della forma (3.2) per qualche $\tilde{E}_i \subset S_i$. Quindi, per l'additività (finito o numerabile) di $\mathbb{P}$ e poi per (3.1), vale

$$
\begin{aligned}
&\mathbb{P}(E_i) = \\
&\sum_{x_1 \in S_1} \cdots \sum_{x_{i-1} \in S_{i-1}} \sum_{x_i \in \tilde{E}_i} \sum_{x_{i+1} \in S_{i+1}} \cdots \sum_{x_n \in S_n} \mathbb{P}\left(\{(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)\}\right) = \\
&\sum_{x_1 \in S_1} \cdots \sum_{x_{i-1} \in S_{i-1}} \sum_{x_i \in \tilde{E}_i} \sum_{x_{i+1} \in S_{i+1}} \cdots \sum_{x_n \in S_n} \prod_{i=1}^{n} \mathbb{P}_i(\{x_i\}) = \\
&\left(\sum_{x_1 \in S_1} \mathbb{P}_1(\{x_1\})\right) \cdots \left(\sum_{x_{i-1} \in S_{i-1}} \mathbb{P}_{i-1}(\{x_{i-1}\})\right) \left(\sum_{x_i \in \tilde{E}_i} \mathbb{P}_i(\{x_i\})\right) \cdot \\
&\left(\sum_{x_{i+1} \in S_{i+1}} \mathbb{P}_{i+1}(\{x_{i+1}\})\right) \cdots \left(\sum_{x_n \in S_n} \mathbb{P}_n(\{x_n\})\right).
\end{aligned}
\tag{3.4}
$$

Nell'ultimo membro tutte le somme tranne la i–esima valgono 1, mentre la i–esima vale $\mathbb{P}_i(\tilde{E}_i)$. Da cui concludiamo che

$$
\mathbb{P}(E_i) = \mathbb{P}_i(\tilde{E}_i). \tag{3.5}
$$

Per concludere procediamo come segue:

$$
\begin{aligned}
&\mathbb{P}\left(\cap_{i=1}^n E_i\right) = \mathbb{P}\left(\tilde{E}_1 \times \tilde{E}_2 \times \cdots \tilde{E}_n\right) = \sum_{x_1 \in \tilde{E}_1} \sum_{x_2 \in \tilde{E}_2} \cdots \sum_{x_n \in \tilde{E}_n} \mathbb{P}\left(\{(x_1, x_2, \dots, x_n)\}\right) = \\
&\sum_{x_1 \in \tilde{E}_1} \sum_{x_2 \in \tilde{E}_2} \cdots \sum_{x_n \in \tilde{E}_n} \mathbb{P}_1(\{x_1\})\mathbb{P}_2(\{x_2\}) \cdots \mathbb{P}_n(\{x_n\}) = \prod_{i=1}^{n} \left(\sum_{x_i \in \tilde{E}_i} \mathbb{P}_i(\{x_i\})\right) = \\
&\prod_{i=1}^{n} \mathbb{P}_i(\tilde{E}_i) = \prod_{i=1}^{n} \mathbb{P}(E_i).
\end{aligned}
\tag{3.6}
$$

Si noti che ciò conclude la dimostrazione di (3.3). Nella suddetta espressione la prima identità segue dall'identità insiemistica

$$
\cap_{i=1}^n E_i = \tilde{E}_1 \times \tilde{E}_2 \times \cdots \times \tilde{E}_n,
$$

che a sua volta segue da (3.2). La seconda identità segue da (3.1), la terza dalla proprietà distributiva somma–prodotto, la quarta dalla l'additività (finita o numerabike) di $\mathbb{P}_i$ e la quinta da (3.5).

Dimostriamo ora il secondo punto della proposizione. Per ipotesi, $\mathbb{P}_i(\{x_i\}) = 1/|S_i|$ per ogni $x_i \in S_i$ e per ogni i. Per (3.1) abbiamo

$$
\mathbb{P}(\{(x_1, x_2 \cdots, x_n)\}) = \mathbb{P}_1(\{x_1\})\mathbb{P}_2(\{x_2\}) \cdots \mathbb{P}_n(\{x_n\}) = \frac{1}{|S_1|} \cdot \frac{1}{|S_2|} \cdots \frac{1}{|S_n|} = \frac{1}{|S|}
$$

Quindi tutti gli eventi elemetari di S hanno probabilità $1/|S|$ e perciò S ha esiti equiprobabili. $\square$

Esempio 1. Consideriamo n prove indipendenti, dove ogni prova ha due esiti che chiamiamo "successo" e "insuccesso", per cui il successo si verifica con probabilità $p \in [0, 1]$. Codificando "successo" e "insuccesso" rispettivamente con le cifre "1" e "0", la singola prova i-esima è modellizzata dallo spazio di probabilità $(S_i, \mathbb{P}_i)$, dove $S_i = \{0, 1\}$ e $\mathbb{P}_i(\{1\}) = p$,

$\mathbb{P}_i(\{0\}) = 1 - p =: q$. Lo spazio prodotto che modellizza l'esperimento globale dato dalle n prove, detto *schema di Bernoulli di parametro* p, è quindi $(S, \mathbb{P})$ dove

$$S = \{0,1\}^n = \{(x_1, x_2, \ldots, x_n) \,:\, x_i \in \{0,1\} \;\forall i : 1 \leq i \leq n\}$$
$$\mathbb{P}(\{(x_1, x_2, \ldots, x_n)\}) = p^k q^{n-k}\,, \qquad k := \sharp\{i \,:\, x_i = 1,\; 1 \leq i \leq n\}$$

Esempio 2. Si lanciano 2 monete truccate per cui testa esce rispettivamente con con probabilità 1/3 e 1/4. Vogliamo determinare la probabilità che escano due facce uguali. Considerando $S = \{(T,T,),(C,C),(T,C),(C,T)\}$, dobbiamo calcolare $\mathbb{P}(E)$ dove $E = \{(T,T),(C,C)\}$. Per l'additività finita e per (3.1) abbiamo

$$\mathbb{P}(E) = \mathbb{P}(\{(T,T)\}) + \mathbb{P}(\{(C,C)\}) = (1/3)(1/4) + (2/3)(3/4) = 7/12\,.$$

4. Variabili aleatorie reali

Proposizione 6. *Data $X : S \to \mathbb{R}$ variabile aleatoria discreta, vale:*

$$Var(X) = 0 \textit{ se e solo se } P(X = E(X)) = 1.$$

Proof. Sia $\{x_i\}$ l'insieme numerabile dei possibili valori assunti da X. Supponiamo prima che $Var(X) = 0$. Per la proposizione che permette di calcolare $E(f(X))$ con $f(u) = (u - E(X))^2$ abbiamo

$$Var(X) = E(f(X)) = \sum_i (x_i - E(X))^2 p_X(x_i)\,. \tag{4.1}$$

Siccome $Var(X) = 0$ tutti gli addendi nel membro destro devono essere nulli, quindi se $p_X(x_i) > 0$ allora $x_i = EX$. Quindi l'unico valore possibile assunto con probabilità positiva è $E(X)$ (deve essere assunto altrimenti sarebbe $p_X(x_i) = 0$ per ogni i mentre sappiamo che $1 = \sum_i p_X(x_i)$). Inoltre, siccome $1 = \sum_i p_X(x_i)$ e siccome $p_X(x_i) = 0$ se $x_i \neq E(X)$, deve valere $p_X(E(X)) = 1$. Abbiamo quindi provato che $1 = p_X(E(X)) = P(X = E(X))$.

Supponiamo viceversa che $P(X = E(X)) = 1$. Ne deriva che $E(X)$ deve essere un possibile valore di X (altrimenti l'evento $\{X = E(X)\}$ sarebbe impossibile e quindi la sua probabilità sarebbe 0). A meno di rinominare i possibili valori, posso assumure $x_1 = E(X)$. Si noti che $P(X = E(X)) = 1$ implica che $P(X \neq E(X)) = 0$. Sia ora x_i un possibile valore di X diverso da $E(X)$. Siccome $\{X = x_i\} \subset \{X \neq E(X)\}$, per monotonia abbiamo

$$P(X = x_i) \leq P(X \neq E(X)) = 0$$

da cui concludiamo che $p_X(x_i) = P(X = x_i) = 0$. Per (4.1) concludiamo che

$$\begin{aligned} Var(X) &= \sum_i (x_i - E(X))^2 p_X(x_i) \\ &= (x_1 - E(X))^2 p_X(x_1) + \sum_{i \neq 1} (x_i - E(X))^2 p_X(x_i) = 0 \cdot 1 + \sum_{i \neq 1} (x_i - E(X))^2 \cdot 0 = 0 \end{aligned}$$

Abbiamo quindi provato che $Var(X) = 0$. □

Proposizione 7. *Sia S numerabile e sia $X : S \to \mathbb{R}$ variabile aleatoria. Allora X è v.a. discreta e $E(X) = \sum_{s \in S} X(s) P(\{s\})$.*

Proof. L'insieme dei possibili valori di X è dato da $X(S)$, immagine di uno spazio numerabile, quindi è esso stesso un insieme numerabile. Ne deriva che X è v.a. discreta.

Sia $\{x_i\} = X(S)$ l'insieme dei possibili valori di X. Abbiamo

$$\sum_{s\in S} X(s)P(\{s\}) = \sum_i \sum_{\substack{s\in S:\\ X(s)=x_i}} X(s)P(\{s\}) = \sum_i \sum_{\substack{s\in S:\\ X(s)=x_i}} x_i P(\{s\}) = \sum_i x_i \sum_{\substack{s\in S:\\ X(s)=x_i}} P(\{s\}) \tag{4.2}$$

Siccome S è numerable e $\{X = x_i\} = \{s \in S : X(s) = x_i\} \subset S$, per additività finita o numerabile abbiamo

$$P(X = x_i) = \sum_{\substack{s\in S:\\ X(s)=x_i}} P(\{s\}).$$

Inserendo la suddetta uguaglianza nell'ultimo membro di (4.2) otteniamo

$$\sum_{s\in S} X(s)P(\{s\}) = \sum_i x_i \sum_{\substack{s\in S:\\ X(s)=x_i}} P(\{s\}) = \sum_i x_i P(X = x_i)$$

e ora basta osservare che l'ultimo membro è per definizione $E(X)$. □

5. Variabili aleatorie discrete vettoriali

Nel seguito, salvo indicazione esplicita, (S, P) sarà un generico spazio di probabilità. Inoltre, "v.a." sarà l'abbreviazione di "variabile aleatoria".

Definizione 2. *Una* **v.a.** m**–dimensionale discreta** *(detta anche* vettore aleatorio discreto*) è una funzione $X = (X_1, X_2, \dots, X_m) : S \to \mathbb{R}^m$ per cui le funzioni $X_1, X_2, \dots, X_m$ sono v.a. discrete.*

Esempio. Segliamo a caso uno studente della classe e descriviamo la sua data di nascita con la terna di numeri (X_1, X_2, X_3) dove X_1 è il numero del giorno di nascita, X_2 il mese usando la numerazione da 1 a 12 (1=gennaio,..) e X_3 è l'anno di nascita.

Una v.a. 1–dimensionale discreta è semplicemente una v.a. discreta (reale, cioè a valori in $\mathbb{R}$).

Proposizione 8. *Se X è v.a. m–dim. discreta allora X può assumere al più un'infinità numerabile di valori.*

Proof. Osserviamo che X può assumere solo valori in

$$W_1 \times W_2 \times \cdots \times W_m = \{(w_1, w_2, \dots, w_m) : w_1 \in W_1,\ w_2 \in W_2\,,\ \dots, w_m \in W_m\}$$

dove in generale W_k è l'insieme dei possibili valori assunti da X_k. Siccome il prodotto cartesiano di insiemi numerabile è un insieme numerabile, otteniamo che $W_1\times W_2\times\cdots\times W_m$ è numerabile e quindi i possibili valori di X formano un insieme numerabile (uso che un sottoinsieme di un insieme numerabile è numerabili). □

Data X v.a. m–dim. discreta, definiamo la **densità discreta di** X come la mappa $p_X : \mathbb{R}^m \to [0, 1]$ tale che

$$p_X(a) = P(X = a), \qquad \forall a \in \mathbb{R}^m.$$

Proposizione 9. *Data X v.a. m–dim. discreta, valgono le seguenti proprietà:*
(i) $p_X(a) \geq 0$ *per ogni* $a \in \mathbb{R}^m$,
(ii) $p_X(a) > 0$ *per un insieme numerabile di valori* $a \in \mathbb{R}^m$,
(iii) $\sum_{a \in \mathbb{R}^m \,:\, p_X(a)>0} p_X(a) = 1$.

Viceversa, data una mappa $p : \mathbb{R}^m \to [0,1]$ *soddisfacente le suddette tre proprietà con "p_X" sostituito da 'p", esiste una v.a. m–dim. discreta X tale che $p = p_X$.*

Si noti che il punto (ii) giustifica la notazione al punto (iii): la somma nel punto (iii) è una somma finita o una serie, oggetti ben definiti. La dimostrazione della suddetta proposizione non fa parte del programma.

Si noti che la suddetta proposizione vale anche per $m = 1$.

Definizione 3. *Se $X_1, \dots, X_m : S \to \mathbb{R}$ sono v.a. discrete, la densità discreta p_X della v.a. m–dim. $X = (X_1, \dots, X_m)$ è detta* **densità congiunta** *delle v.a. $X_1, \dots, X_m$, mentre le densità discrete p_{X_1}, ..., p_{X_m} rispettivamente di $X_1, \dots, X_m$ sono dette* **densità marginali** *di X.*

Proposizione 10. *Nota la densità congiunta si ottengono le densità marginali. In generale, date $X_1, \dots, X_m : S \to \mathbb{R}$ v.a. discrete, la densità marginale p_{X_i} si ottiene come*

$$p_{X_i}(a) = \sum_{a_1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_m} P_{X_1, X_2, \dots, X_m}(a_1, a_2, \dots, a_{i-1}, a, a_{i+1}, \dots, a_m)\,. \qquad (5.1)$$

dove a_1 varia tra i possibili valori di X_1, a_2 varia tra i possibili valori di X_2 e così via. Per esempio nel caso $m = 2$ vale

$$p_{X_1}(a) = \sum_{a_2} p_{X_1,X_2}(a, a_2)\,, \qquad p_{X_2}(a) = \sum_{a_1} p_{X_1,X_2}(a_1, a)\,, \qquad (5.2)$$

dove nella prima somma a_2 varia tra i possibili valori di X_2 e nella seconda somma X_1 varia tra i possibili valori di X_1.

Infine, le densità marginali non determinano la densità congiunta.

Proof. Consideriamo per semplicità di notazione il caso $m = 2$, per cui (5.1) si riduce a (5.2). Chiamo W_2 l'insieme dei possibili valori di X_2. Allora

$$\begin{aligned} p_{X_1}(a) &= P(X_1 = a) = P\left(X_1 = a\,,\, X_2 \in W_2\right) \\ &= \sum_{a_2 \in W_2} P\left(X_1 = a,\ X_2 = a_2\right) = \sum_{a_2 \in W_2} p_{X_1,X_2}(a, a_2) \end{aligned}$$

(si noti che la terza identità nella suddetta espressione segue dalla additivà - numerabile o finita - di P). In conclusione, $p_{X_1}(a) = \sum_{a_2 \in W_2} p_{X_1,X_2}(a, a_2)$. Analogamente si prova che $p_{X_2}(a) = \sum_{a_1 \in W_1} p_{X_1,X_2}(a_1, a)$ dove W_1 è l'insieme dei possibili valori di X_1. Questo conclude la dimostrazione di (5.2).

Per dimostrare che le densità marginali non determinano quella congiunta, diamo il seguente controesempio. Lancio due volte una moneta e, dato $i = 1, 2$, chiamo U_i la v.a. che vale 1 se al lancio i–esimo esce testa e 0 se al lancio i–esimo viene croce. Pongo $X_1 := U_1$, $X_2 := U_2$, $X_1' := U_1$ e $X_2' := U_1$. È facile verificare che $p_{X_1} = p_{X_1'}$ e $p_{X_2} = p_{X_2'}$, mentre $p_{X_1,X_2} \neq p_{X_1',X_2'}$ (lasciamo la verifica come facile esercizio). $\square$

• **Esperimento 1**. Da un'urna contenente 6 palline numerate da 1 a 6 si estraggono due palline con rimpazzo. Indichiamo con X_1 e X_2 i risultati delle due estrazioni, e poniamo

$X = (X_1, X_2)$. L'esperimento è descritto dallo spazio campionario

$$S = \big\{(u,v) \,:\, u,v \in \{1,\dots,6\}\big\}\,,$$

avente esiti equiprobabili. Si noti che $X_1(u,v) = u$, $X_2(u,v) = v$ e $X(u,v) = (u,v)$. È facile calcolare p_{X_1}, p_{X_2} e p_X:

$$p_{X_1}(c) = p_{X_2}(c) = \begin{cases} 1/6 & \text{se } c \in \{1,\dots,6\}, \\ 0 & \text{altrimenti,} \end{cases}$$

mentre

$$p_X(u,v) = \begin{cases} 1/36 & \text{se } (u,v) \in S, \\ 0 & \text{altrimenti.} \end{cases}$$

• **Esperimento 2**. Dalla suddetta urna si estraggono due palline senza rimpiazzo. Indichiamo con X_1' e X_2' i risultati delle due estrazioni, e poniamo $X = (X_1', X_2')$. L'esperimento è descritto dallo spazio campionario

$$S' = \big\{(u,v) \,:\, u,v \in \{1,\dots,6\},\, u \neq v\big\}$$

avente esiti equiprobabili. Si noti che $X_1'(u,v) = u$, $X_2'(u,v) = v$ e $X'(u,v) = (u,v)$. È facile calcolare $p_{X_1'}$, $p_{X_2'}$ e $p_{X'}$:

$$p_{X_1'}(c) = p_{X_2'}(c) = \begin{cases} 1/6 & \text{se } c \in \{1,\dots,6\}, \\ 0 & \text{altrimenti,} \end{cases}$$

mentre

$$p_{X'}(u,v) = \begin{cases} 1/30 & \text{se } (u,v) \in S', \\ 0 & \text{altrimenti.} \end{cases}$$

• Confrontando i due esperimenti, otteniamo che $p_{X_1} = p_{X_1'}$, $p_{X_2} = p_{X_2'}$ ma $p_X \neq p_{X'}$, quindi le densità marginali non determinano univocamente la densità congiunta come già dimostrato.

6. Indipendenza di variabili aleatorie

Definizione. *Le v.a.* $X_1, X_2, \dots, X_m : S \to \mathbb{R}$ *(definite sullo stesso spazio di probabilità) si dicono indipendenti se per ogni scelta di* $A_1, A_2, \dots, A_m \subset \mathbb{R}$ *si ha*

$$P\left(X_1 \in A_1,\, X_2 \in A_2, \dots, X_m \in A_m\right) = P(X_1 \in A_1)P(X_2 \in A_2)\cdots P(X_m \in A_m)\,. \tag{6.1}$$

Si noti che la definizione vale in generale, anche per v.a. non discrete.

Proposizione 11. *Le v.a.* $X_1, \dots X_m : S \to \mathbb{R}$ *sono indipendenti se e solo se per ogni scelta di* $A_1, A_2, \dots, A_m \subset \mathbb{R}$ *gli eventi*

$$\big\{X_1 \in A_1\big\},\, \big\{X_2 \in A_2\big\}, \dots, \big\{X_m \in A_m\big\}$$

sono indipendenti.

Dim. Supponiamo $X_1, \dots, X_m$ essere indipendenti e fissiamo $A_1, A_2, \dots, A_m \subset \mathbb{R}$. Per provare che gli eventi $\big\{X_1 \in A_1\big\}$, $\big\{X_2 \in A_2\big\}, \dots, \big\{X_m \in A_m\big\}$ sono indipendenti dobbiamo provare che

$$P\left(\{X_{i_1} \in A_{i_1}\} \cap \dots \cap \{X_{i_r} \in A_{i_r}\}\right) = P\left(X_{i_1} \in A_{i_1}\right) \cdots P\left(X_{i_r} \in A_{i_r}\right) \tag{6.2}$$

per ogni scelta di indici $1 \le i_1 < i_2 < \cdots < i_r \le m$ are variare di $r = 2, 3, ..., m$. Siccome $\{X_{i_1} \in A_{i_1}\} \cap \cdots \cap \{X_{i_r} \in A_{i_r}\} = \cap_{i=1}^m \{X_i \in A'_i\}$ dove

$$A'_i = \begin{cases} A_i & \text{se } i \in \{i_1, \ldots, i_r\} \\ \mathbb{R} & \text{altrimenti}\,, \end{cases}$$

per la definizione di indipendenza abbiamo

$$P\left(\{X_{i_1} \in A_{i_1}\} \cap \cdots \cap \{X_{i_r} \in A_{i_r}\}\right) = \prod_{i=1}^m P(X_i \in A'_i).$$

Usando che

$$P(X_i \in A'_i) = \begin{cases} P(X_i \in A_i) & \text{se } i \in \{i_1, \ldots, i_r\} \\ 1 & \text{altrimenti} \end{cases}$$

otteniamo (6.2).

Viceversa, supponiamo che per ogni per ogni scelta di $A_1, A_2, \ldots, A_m \subset \mathbb{R}$ gli eventi $\big\{X_1 \in A_1\big\}$, $\big\{X_2 \in A_2\big\}, \ldots, \big\{X_m \in A_m\big\}$ sono indipendenti. Allora (6.1) è banalmente verificata, da cui l'indipendenza di $X_1, \ldots, X_m$. $\square$

Diamo ora un criterio molto utile per verificare l'indipendenza di v.a. discrete:

Proposizione 12. *Siano $X_1, \ldots X_m : S \to \mathbb{R}$ v.a. discrete. Allora $X_1, \ldots, X_m$ sono indipendenti se e solo se*

$$p_{X_1, X_2, \ldots, X_m}(a_1, a_2, \ldots, a_m) = p_{X_1}(a_1) p_{X_2}(a_2) \cdots p_{X_m}(a_m) \tag{6.3}$$

dove $a_1, a_2, \ldots, a_m$ variano tra i possibili valori di $X_1, X_2, \ldots, X_m$, rispettivamente.

Dim. Per semplicità di notazione consideriamo solo il caso $m = 2$, il caso generale si tratta con gli stessi ragionamenti. Supponiamo dapprima che X_1, X_2 siano v.a. discrete indipendenti e proviamo (6.3). Per definizione,

$$p_{X_1, X_2}(a_1, a_2) = P(X_1 = a, X_2 = a_2)\,.$$

Per l'indipendenza di X_1, X_2 , l'ultimo membro può essere scritto come $P\big(X_1 = a_1\big)P\big(X_2 = a_2\big)$. Quindi possiamo concludere che

$$p_{X_1, X_2}(a_1, a_2) = P(X_1 = a, X_2 = a_2) = P\big(X_1 = a_1\big)P\big(X_2 = a_2\big) = p_{X_1}(a_1) p_{X_2}(a_2)\,.$$

Supponiamo ora che valga (6.3) con $m = 2$, per ogni a_1, a_2 possibili valori di X_1, X_2 rispettivamente, e proviamo l'indipendenza di X_1 e X_2. Denotiamo con W_1, W_2 l'insieme dei possibili valori di X_1, X_2 rispettivamente. Fissati $A_1, A_2 \subset \mathbb{R}$ possiamo scrivere

$$P(X_1 \in A_1,\, X_2 \in A_2) = P\left(\cup_{a_1 \in A_1 \cap W_1,\, a_2 \in A_2 \cap W_2} \{X_1 = a_1, X_2 = a_2\}\right)\,. \tag{6.4}$$

Per l'additività di P (finita o numerabile),

$$P\left(\cup_{a_1 \in A_1 \cap W_1, a_2 \in A_2 \cap W_2} \{X_1 = a_1, X_2 = a_2\}\right) = \sum_{a_1 \in A_1 \cap W_1} \sum_{a_2 \in A_2 \cap W_2} P(X_1 = a_1, X_2 = a_2)\,. \tag{6.5}$$

Per (6.3) con $m = 2$ abbiamo che

$$P(X_1 = a_1, X_2 = a_2) = p_{X_1, X_2}(a_1, a_2) = p_{X_1}(a_1) p_{X_2}(a_2) = P(X_1 = a_1) P(X_2 = a_2).$$

Quindi abbiamo

$$\sum_{a_1\in A_1\cap W_1}\sum_{a_2\in A_2\cap W_2} P(X_1=a_1,X_2=a_2)=$$
$$\sum_{a_1\in A_1\cap W_1}\sum_{a_2\in A_2\cap W_2} P(X_1=a_1)P(X_2=a_2)=$$
$$\left[\sum_{a_1\in A_1\cap W_1} P(X_1=a_1)\right]\cdot\left[\sum_{a_2\in A_2\cap W_2} P(X_2=a_2)\right]=$$
$$P(X_1\in A_1\cap W_1)P(X_2\in A_2\cap W_2)=P(X_1\in A_1)P(X_2\in A_2)\,. \tag{6.6}$$

Combinando (6.4), (6.5) e (6.6) otteniamo che

$$P(X_1\in A_1,X_2\in A_2)=P(X_1\in A_1)P(X_2\in A_2)$$

quindi X_1,X_2 sono indipendenti. □

Esempi. Applicando il suddetto criterio o verificando le ipotesi nella definizione di indipendenza è facile verificare che le variabili X_1,X_2 dell'esperimento 1 sono indipendenti, mentre le variabili X_1',X_2' dell'esperimento 2 non sono indipendenti.

Nota. Grazie alla Proposizione 2, otteniamo che, se $X_1,\dots,X_m:S\to\mathbb{R}$ sono v.a. discrete indipendenti, allora la densità congiunta di $X_1,\dots,X_m$ è univocalmente determinata dalle densità marginali (infatti vale (6.7)).

Per v.a. in generale, non necessariamente discrete, ricordiamo che vale il seguente criterio di cui non richiediamo la dimostrazione:

Proposizione 13. *Siano $X_1,\dots X_m:S\to\mathbb{R}$ v.a. Allora, $X_1,\dots,X_m$ sono indipendenti se e solo se*

$$F_{X_1,\dots,X_m}(a_1,\dots,a_m)=F_{X_1}(a_1)\cdots F_{X_m}(a_m) \tag{6.7}$$

per ogni $a_1,\dots,a_m\in\mathbb{R}$, dove la funzione di distribuzione congiunta $F_{X_1,\dots,X_m}:\mathbb{R}^m\to[0,1]$ è definita come $F_{X_1,\dots,X_m}(a_1,\dots,a_m)=P(X_1\le a_1,\dots,X_m\le a_m)$, mentre F_{X_i} è la funzione di distribuzione di X_i, ovvero $F_{X_i}(a)=P(X_i\le a)$ per ogni $a\in\mathbb{R}$.

7. Valor medio della funzione di una v.a. m–dimensionale

La seguente proposizione estende un risultato già studiato nel caso $m=1$, la relativa dimostrazione è completamente analoga al caso $m=1$ e quindi verrà omessa:

Proposizione 14. *Siano $X_1,\dots,X_m:S\to\mathbb{R}$ v.a. discrete, sia $X=(X_1,\dots,X_m):S\to\mathbb{R}^m$ e sia $f:\mathbb{R}^m\to\mathbb{R}$ una generica funzione. Allora la mappa composta $f(X)=f\circ X:S\to\mathbb{R}$ è una v.a. discreta. Inoltre,*

$$\mathbf{E}\left(f(X)\right)=\sum_{a_1,a_2,\dots,a_m} f\left(a_1,a_2,\dots,a_m\right)p_{X_1,X_2,\dots,X_m}\big(a_1,a_2,\dots,a_m\big)\,, \tag{7.1}$$

dove $a_1,a_2,\dots,a_m$ variano tra i possibili valori di $X_1,X_2,\dots,X_m$ rispettivamente.

Esempio. Consideriamo le v.a. X_1' e X_2' dell'esperimento 2. Vogliamo calcolare $\mathbf{E}(X_1'\cdot X_2')$. A tal fine possiamo usare la Prop. 14 dato che $X_1'\cdot X_2'=f(X')$ dove $X'=(X_1',X_2')$

e $f:\mathbb{R}^2\to\mathbb{R}$ è definita come $f(u,v)=u\cdot v$. Quindi

$$\mathbf{E}(X_1'\cdot X_2')=\sum_{u=1}^{6}\sum_{v=1}^{6}u\cdot v\mathbf{P}(X_1'=u,X_2'=v)=\frac{1}{30}\sum_{u,v:1\le u,v\le 6,\ u\neq v}u\cdot v=$$

$$\frac{1}{30}\left[\sum_{u=1}^{6}\sum_{v=1}^{6}u\cdot v-1^2-2^2-\cdots-6^2\right]=\frac{1}{30}\left[\Big(\sum_{u=1}^{6}u\Big)^2-1^2-2^2-\cdots-6^2\right]=...$$

Indichiamo due utili corollari della precedente proposizione:

Corollario 1. *Se $X_1,\dots,X_m:S\to\mathbb{R}$ sono v.a. discrete e $c_1,\dots,c_m\in\mathbb{R}$, allora vale*

$$\mathbf{E}(c_1X_1+\cdots+c_mX_m)=c_1\mathbf{E}(X_1)+\cdots+c_m\mathbf{E}(X_m)\,. \tag{7.2}$$

Dim. Per semplicità di notazione, supponiamo $m=2$. Il caso generale è completamente analogo. Grazie alla suddetta proposizione 14 con $f(u,v)=c_1u+c_2v$, otteniamo che

$$\begin{aligned}
&\mathbf{E}(c_1X_1+c_2X_2)=\sum_{a_1,a_2}\big(c_1a_1+c_2a_2)P\left(X_1=a_1,X_2=a_2\right)=\\
&c_1\sum_{a_1,a_2}a_1P\left(X_1=a_1,X_2=a_2\right)+c_2\sum_{a_1,a_2}a_2P\left(X_1=a_1,X_2=a_2\right)=\\
&c_1\sum_{a_1}a_1\left(\sum_{a_2}P\left(X_1=a_1,X_2=a_2\right)\right)+c_2\sum_{a_2}a_2\left(\sum_{a_1}P\left(X_1=a_1,X_2=a_2\right)\right)=\\
&c_1\sum_{a_1}a_1P(X_1=a_1)+c_2\sum_{a_2}a_2P(X_2=a_2)=c_1E(X_1)+c_2E(X_2)\,.
\end{aligned} \tag{7.3}$$

Sopra a_1,a_2 variano tra i possibili valori di X_1,X_2, rispettivamente. □

Corollario 2. *Se $X_1,X_2,\dots,X_m:S\to\mathbb{R}$ sono v.a. discrete* **indipendenti**, *allora*

$$\mathbf{E}(X_1\cdot X_2\cdots X_m)=\mathbf{E}(X_1)\cdot\mathbf{E}(X_2)\cdots\mathbf{E}(X_m)\,.$$

Dim. Per semplicità di notazione, supponiamo $m=2$. Il caso generale è completamente analogo. Grazie alla Proposizione 14 otteniamo che

$$\begin{aligned}
\mathbf{E}(X_1\cdot X_2)&=\sum_{a_1,a_2}a_1\cdot a_2\cdot P\left(X_1=a_1,\ X_2=a_2\right)=\\
&=\sum_{a_1,a_2}a_1\cdot a_2\cdot P\left(X_1=a_1\right)P(X_2=a_2)=\\
&\left[\sum_{a_1}a_1P(X_1=a_1)\right]\cdot\left[\sum_{a_2}a_2P(X_2=a_2)\right]=E(X_1)E(X_2)\,.
\end{aligned}$$

Si noti che nella seconda identità abbiamo usato l'indipendenza di X_1,X_2. □

Proposizione 15. *Siano $X_1,\dots,X_n$ variabili aleatorie indipendenti (quindi definite sullo stesso spazio di probabilità). Allora date le funzioni $f_1,f_2,\dots,f_n:\mathbb{R}\to\mathbb{R}$, le variabili aleatorie $f_1(X_1),\dots,f_n(X_n)$ sono indipendenti ed in particolare vale*

$$\mathbf{E}\left[\prod_{i=1}^{n}f_i(X_i)\right]=\prod_{i=1}^{n}\mathbf{E}\big[f_i(X_i)\big] \tag{7.4}$$

Proof. Dimostriamo prima l'indipendenza, ovvero dobbiamo provare che per ogni famiglia di insiemi $A_1, A_2, \ldots, A_n \subset \mathbb{R}$ vale

$$\mathbf{P}(f_i(X_i) \in A_i \ \forall i : 1 \leq i \leq n) = \prod_{i=1}^{n} \mathbf{P}(f_i(X_i) \in A_i) .$$

Dato che $f_i(X_i) \in A_i$ se e solo se $X_i \in f_i^{-1}(A_i)$ possiamo scrivere

$$\mathbf{P}(f_i(X_i) \in A_i \ \forall i : 1 \leq i \leq n) = \mathbf{P}\left(X_i \in f_i^{-1}(A_i) \ \forall i : 1 \leq i \leq n\right) .$$

Per l'indipendenza di $X_1, X_2, \ldots, X_n$ possiamo riscrivere il suddetto membro destro come

$$\mathbf{P}\left(X_i \in f_i^{-1}(A_i) \ \forall i : 1 \leq i \leq n\right) = \prod_{i=1}^{n} \mathbf{P}\left(X_i \in f_i^{-1}(A_i)\right)$$

Usando di nuovo che $f_i(X_i) \in A_i$ se e solo se $X_i \in f_i^{-1}(A_i)$ possiamo scrivere

$$\prod_{i=1}^{n} \mathbf{P}\left(X_i \in f_i^{-1}(A_i)\right) = \mathbf{P}\left(X_i \in f_i^{-1}(A_i) \ \forall i : 1 \leq i \leq n\right) .$$

Mettendo insieme le suddette uguaglianza otteniamo l'indipendenza.

Nel caso di v.a. discrete indipendenti abbiamo provato che il valore atteso del prodotto coincide con il prodotto dei singoli valori attesi (si veda il file di integrazioni). Tale propietà vale anche per v.a.continue, e anche miste... ma non richiedo la dimostrazione nel caso generale. Usando tale propietà e l'indipendenza appena provata di $f_1(X_1), \ldots, f_n(X_n)$ otteniamo immediatamente (7.4). □

8. Varianza di una somma di variabili aleatorie

Definizione 4. *Date $X_1, X_2 : S \to \mathbb{R}$ v.a. discrete, definiamo la covarianza $Cov(X_1, X_2)$ tra X_1 e X_2 come*

$$Cov(X_1, X_2) = \mathbf{E}\left[(X_1 - \mathbf{E}(X_1))(X_2 - \mathbf{E}(X_2))\right] . \tag{8.1}$$

Se $\mathrm{Cov}(X_1, X_2) = 0$ *diciamo che X_1, X_2 sono* **non correlate***.*

Proposizione 16. *Vale $Cov(X_1, X_2) = Cov(X_2, X_1)$, $Cov(X, X) = Var(X)$ e*

$$Cov(X_1, X_2) = \mathbf{E}(X_1 X_2) - \mathbf{E}(X_1)\mathbf{E}(X_2) \tag{8.2}$$

Dim. Le prime due uguaglianza sono banali. Proviamo (8.2). Sviluppando il prodotto otteniamo che

$$\mathbf{E}\left[(X_1 - \mathbf{E}(X_1))(X_2 - \mathbf{E}(X_2))\right] = \mathbf{E}\left[X_1 X_2 - X_1 \mathbf{E}(X_2) - \mathbf{E}(X_1) X_2 + \mathbf{E}(X_1)\mathbf{E}(X_2)\right] .$$

Grazie alla linearità dell'operatore valor medio, possiamo riscrivere il membro destro della suddetta espressione come

$$\mathbf{E}(X_1 X_2) - \mathbf{E}(X_1)\mathbf{E}(X_2) - \mathbf{E}(X_1)\mathbf{E}(X_2) + \mathbf{E}(X_1)\mathbf{E}(X_2) = \mathbf{E}(X_1 X_2) - \mathbf{E}(X_1)\mathbf{E}(X_2) ,$$

da cui (8.2). □

Proposizione 17. *Siano $X_1, X_2 : S \to \mathbb{R}$ v.a. discrete. Se X_1, X_2 sono indipendenti allora $Cov(X_1, X_2) = 0$. Il viceversa non è vero.*

Dim. Se X_1, X_2 sono indipendenti, otteniamo che $\mathrm{Cov}(X_1, X_2) = 0$ grazie al Corollario 2. Per provare che il viceversa non è vero basta esibire un controesempio. A tal fine, consideriamo due v.a. X_1, X_2 definite sullo stesso spazio di probabilità a valori $-1, 0, 1$ tali

$$P(X_1 = -1, X_2 = 0) = P(X_1 = 1, X_2 = 0) = \\ P(X_1 = 0, X_2 = -1) = P(X_1 = 0, X_2 = 1) = 1/4\,.$$

Una tale coppia di v.a. esiste grazie alla Proposizione 1. Si noti che

$$\begin{cases} P(X_1 = 1) = P(X_1 = -1) = P(X_2 = 1) = P(X_2 = -1) = 1/4, \\ P(X_1 = 0) = P(X_2 = 0) = 1/2. \end{cases}$$

quindi $\mathbf{E}(X_1) = \mathbf{E}(X_2) = 0$. Siccome $P(X_1X_2 = 0) = 1$ abbiamo anche che $\mathbf{E}(X_1X_2) = 0$. Quindi

$$Cov(X_1, X_2) = \mathbf{E}(X_1X_2) - \mathbf{E}(X_1)\mathbf{E}(X_2) = 0.$$

Banalmente X_1, X_2 non sono indipendenti, infatti

$$P(X_1 = 1, X_2 = 1) = 0,$$

mentre

$$P(X_1 = 1)P(X_2 = 1) = 1/4 \cdot 1/4 > 0$$

e ciò contraddice (6.1).

□

Il concetto di covarianza risulta utile per studiare la varianza di una somma di v.a. discrete definite sullo stesso spazio di probabilità:

Proposizione 18. *Se $X_1, \dots, X_m : S \to \mathbb{R}$ sono v.a. discrete, allora*

$$Var(X_1 + \cdots + X_m) = \sum_{j=1}^{m} Var(X_j) + 2 \sum_{1 \le i < j \le m} Cov(X_i, X_j)\,. \tag{8.3}$$

In particolare, se $X_1, \dots, X_m$ sono indipendenti, o almeno a 2 a 2 indipendenti, o almeno a 2 a 2 non correlate, allora

$$Var(X_1 + \cdots + X_m) = \sum_{j=1}^{m} Var(X_j)\,. \tag{8.4}$$

Dim. (8.4) deriva da (8.3) dato che nei casi considerati vale $\mathrm{Cov}(X_i, X_j) = 0$ con $i \neq j$.

Dimostriamo (8.3). Per definizione di varianza ho:

$$\mathrm{Var}(X_1 + \cdots + X_m) = E\left(\left[\sum_{i=1}^{m} X_i - E(\sum_{i=1}^{m} X_i)\right]^2\right)\,. \tag{8.5}$$

Dato che $E(\sum_{i=1}^{m} X_i) = \sum_{i=1}^{m} E(X_i)$ posso scrivere

$$E\left(\left[\sum_{i=1}^{m} X_i - E(\sum_{i=1}^{m} X_i)\right]^2\right) = E\left(\left[\sum_{i=1}^{m} \{X_i - E(X_i)\}\right]^2\right)\,. \tag{8.6}$$

Osservo che

$$\left[\sum_{i=1}^{m} \{X_i - E(X_i)\}\right]^2 = \sum_{i=1}^{m} \{X_i - E(X_i)\}^2 + 2 \sum_{1 \le i < j \le m} \{X_i - E(X_i)\}\{X_j - E(X_j)\}$$

Quindi per la linerità di $E(\cdot)$ ho

$$E\left(\left[\sum_{i=1}^{m}\{X_i - E(X_i)\}\right]^2\right) =$$

$$\sum_{i=1}^{m} E\left(\{X_i - E(X_i)\}^2\right) + 2\sum_{1\le i<j\le m} E\left(\{X_i - E(X_i)\}\{X_j - E(X_j)\}\right)\,. \quad (8.7)$$

Combinando (8.5), (8.6) e (8.7) ed infine usando la definizione di varianza e covarianza ho

$$\begin{aligned}\mathrm{Var}(X_1 + \cdots + X_m) &= \sum_{i=1}^{m} E\left(\{X_i - E(X_i)\}^2\right) + 2\sum_{1\le i<j\le m} E\left(\{X_i - E(X_i)\}\{X_j - E(X_j)\}\right)\\ &= \sum_{i=1}^{m}\mathrm{Var}(X_i) + 2\sum_{1\le i<j\le m}\mathrm{Cov}(X_i, X_j)\,.\end{aligned}$$

□

Concludiamo con la bilinerità della covarianza:

Proposizione 19. *Siano $X_1, \dots, X_m, Y_1, \dots, Y_n$ v.a. definite sullo stesso spazio di probabilità e siano $a_1, \dots, a_m, b_1, \dots, b_n$ numeri reali. Allora*

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \mathrm{Cov}\left(X_i, Y_j\right)\,. \quad (8.8)$$

Dim. Sia Y v.a. generica definita sullo stesso spazio di probabilità delle $X_1, \dots, X_m$. Allora

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, Y\right) = \mathbf{E}\left([\sum_{i=1}^{m} a_i X_i - \mathbf{E}(\sum_{i=1}^{m} a_i X_i)][Y - \mathbf{E}(Y)]\right) =$$

$$\mathbf{E}\left(\sum_{i=1}^{m} a_i [X_i - \mathbf{E}(X_i)][Y - \mathbf{E}(Y)]\right) = \sum_{i=1}^{m} a_i \mathbf{E}\left([X_i - \mathbf{E}(X_i)][Y - \mathbf{E}(Y)]\right) = \sum_{i=1}^{m} a_i \mathrm{Cov}(X_i, Y)\,,$$

dove la prima e la quarta uguaglianza seguono dalla definizione di covarianza, la seconda e la terza dalla linearità del valore atteso.

Abbiamo provato che $\mathrm{Cov}(\cdot,\cdot)$ è lineare nella prima entrata, analogamente lo è nella seconda. Quindi, applicando prima la linearità nella prima entrata e poi la linearità nella seconda entrata abbiamo

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m} a_i \mathrm{Cov}\left(X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \mathrm{Cov}\left(X_i, Y_j\right)\,.$$

□

9. Integrazione su v.a. di Poisson e v.a. geometrica

La v.a. geometrica gode della seguente proprietà detta *assenza di memoria*:

Proposizione 20. *Se X è v.a. geometrica di parametro p, allora dati n_1, n_2 interi positivi vale*

$$P(X = n_1 + n_2 | X > n_1) = P(X = n_2)\,.$$

Dim.

$$\begin{aligned} P(X = n_1 + n_2 | X > n_1) &= \frac{P(X = n_1 + n_2,\, X > n_1)}{P(X > n_1)} = \frac{P(X = n_1 + n_2)}{P(X > n_1)} \\ &= \frac{p(1-p)^{n_1+n_2-1}}{(1-p)^{n_1}} = p(1-p)^{n_2-1} = P(X = n_2)\,. \end{aligned}$$

□

Proposizione 21. *Siano X, Y v.a. di Possoin indipendenti rispettivamente di parametro λ e μ, definite sullo stesso spazio di probabilità. Allora $X + Y$ è v.a. di Poisson di parametro $\lambda + \mu$.*

Dim. $X + Y$ assume con probabilità positiva solo valori interi non negativi. Sia k un numero intero non negativo. Allora

$$\begin{aligned} P(X + Y = k) &= \sum_{j=0}^{k} P(X = j, Y = k - j) = \sum_{j=0}^{k} e^{-\lambda} \frac{\lambda^j}{j!} e^{-\mu} \frac{\mu^{k-j}}{(k-j)!} = \\ & e^{-(\lambda+\mu)} \sum_{j=0}^{k} \frac{\lambda^j}{j!} \frac{\mu^{k-j}}{(k-j)!} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^{k} \binom{k}{j} \lambda^j \mu^{k-j} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!} \end{aligned}$$

(l'ultima uguaglianza segue dal teorema del binomio). Siccome l'ultima espressione coincide con la probabilità per una v.a. di Poisson di parametro $\lambda + \mu$ di essere pari a k, abbiamo la tesi. □

La v.a. di Poisson, in certe condizioni, approssima la v.a. binomiale. Infatti, vale la seguente proprietà detta anche *legge dei piccoli numeri*:

Proposizione 22. *Per ogni $n \in \{1, 2, 3, \dots\}$ sia $p_n \in (0, 1)$ tale che il limite $\lim_{n\uparrow\infty} np_n$ esiste finito e positivo. Allora, posto $X_n = Bin(n, p_n)$ e $Y = Poisson(\lambda)$ dove $\lambda = \lim_{n\uparrow\infty} np_n$, vale*

$$\lim_{n\uparrow\infty} P(X_n = k) = P(Y = k)\,, \qquad \forall k \in \{0, 1, 2, \dots\}\,.$$

Dim. Dobbiamo dimostrare che per ogni k intero non negativo vale

$$\lim_{n\uparrow\infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}\,.$$

A tal fine osserviamo che

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

e quindi

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{(n\,p_n)^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{(1-p_n)^n}{(1-p_n)^k}\,.$$

Siccome $n\,p_n \to \lambda$ abbiamo che $(n\,p_n)^k \to \lambda^k$. Inoltre, vale

$$\lim_{n\uparrow\infty}\left(1-\frac{1}{n}\right) = \lim_{n\uparrow\infty}\left(1-\frac{2}{n}\right) = \cdots = \lim_{n\uparrow\infty}\left(1-\frac{k-1}{n}\right) = 1\,.$$

Siccome $n\,p_n \to \lambda$ abbiamo che $p_n \to 0$. Quindi $(1-p_n)^k \to 1$. Infine, abbiamo che

$$\lim_{n\uparrow\infty}(1-p_n)^n = \lim_{n\uparrow\infty} e^{n\log(1-p_n)} = e^{\lim_{n\uparrow\infty}(np_n)\frac{\log(1-p_n)}{p_n}}\,.$$

L'ultimo limite si calcola ricordando che $np_n \to \lambda$, mentre per De L'Hospital vale

$$\lim_{x\downarrow 0}\frac{\log(1-x)}{x} = -\lim_{x\downarrow 0}\frac{1}{1-x} = -1\,.$$

Quindi

$$e^{\lim_{n\uparrow\infty}(np_n)\frac{\log(1-p_n)}{p_n}} = e^{-\lambda}\,.$$

Mettendo insieme tutti i limiti calcolati finora otteniamo la tesi. □

10. Valore atteso condizionato

Definizione 5. *Dato un evento F di probabilità positiva e data una v.a. X , definitiamo il valore atteso di X condizionato a F, ovvero $\mathbf{E}[X|F]$, come il valore atteso di X rispetto alla probabilità condizionata $P(\cdot|F)$.*

Proposizione 23. *Siano $F_1, F_2, \ldots, F_n$ eventi di probabilità positiva a due a due disgiunt la cui unione è tutto lo spazio campionario S. Allora data una v.a. X vale*

$$\mathbf{E}(X) = \sum_{i=1}^n \mathbf{E}[X|F_i]P(F_i)\,.$$

Proof. Diamo la dimostrazione assumendo che X sia v.a. discreta. Usando la legge della probabilità totale e scambiando l'ordine delle somme otteniamo:

$$\mathbf{E}(X) = \sum_{x\in\mathbb{R}} xP(X=x) = \sum_{x\in\mathbb{R}} x\sum_{i=1}^n P(X=x|F_i)P(F_i) =$$

$$\sum_{i=1}^n\left(\sum_{x\in\mathbb{R}} xP(X=x|F_i)\right)P(F_i) = \sum_{i=1}^n \mathbf{E}[X|F_i]P(F_i)\,.$$

Infatti: la prima "=" deriva dalla def. di valore atteso, la seconda "=" deriva dalla legge della probabilità totale, la terza "=" segue da uno scambio di somme, l'ultima "=" deriva dalla def. di valore atteso condizionato.

□

Esempio. Vi sono due urne con palline. Nell'urna A vi sono 4B, 3R,2V (4 palline bianche, 3 rosse..). Nell'urna B vi sono 5B, 2R, 1V. Lancio un dado. Se esce 1 estraggo 4 palline dall'urna A, altrimenti estraggo 2 palline dall'urna B. Determinare $\mathbf{E}(X)$ dove X è il numero di palline bianche estratte.
Svolgimento. Sia $F =$ "esce 1". Per la suddetta prop. applicata a F, F^c abbiamo

$$\mathbf{E}(X) = \mathbf{E}(X|F)P(F) + \mathbf{E}(X|F^c)P(F^c) = \mathbf{E}(X|F)(1/6) + \mathbf{E}(X|F^c)(5/6)\,.$$

Rispetto a $P(\cdot|F)$ X è v.a. ipergeometrica(N=9,n=4,m=4). Quindi $\mathbf{E}(X|F) = 4 \cdot 4/9 = 16/9$. Rispetto a $P(\cdot|F^c)$ X è v.a. ipergeometrica(N=8,n=2,m=5). Quindi $\mathbf{E}(X|F) = 2 \cdot 5/8 = 5/4$. Mettendo insieme tutti i pezzi abbiamo $\mathbf{E}(X) = (16/9)(1/6) + (5/4)(5/6)$.

11. Legge di una v.a.

Proposizione 24. *Siano date due v.a. X, Y tali che in un opportuno intorno dell'origine valga $M_X(t) = M_Y(t) < \infty$. Allora X e Y hanno la stessa distribuzione.*

Ricordiamo la seguente definizione:

Definizione 6. *La legge, o distribuzione, di una v.a. X è la mappa*

$$\mathbb{R} \supset A \to \mathbf{P}(X \in A) \in [0,1] .$$

Ad essere precisi nella suddetta definizione uno dovrebbe restringersi ai sottinsiemi A borelliani, ma tralasciamo questi dettagli di teoria della misura.

Proposizione 25. *Siano X, Y v.a. discrete. Allora X, Y hanno la stessa distribuzione se e solo se $p_X = p_Y$ (uguaglianza delle densità discrete).*
Siano X, Y v.a. continue. Allora X, Y hanno la stessa distribuzione se e solo se $f_X(z) = f_Y(z)$ per tutti i valori $z \in \mathbb{R}$ a meno di un insieme di lunghezza 0 (uguaglianza delle funzioni di densità a meno di insieme di lunghezza 0).

Proof. Supponiamo X, Y siano v.a. discrete aventi la stessa legge. Prendendo $A = \{z\}$ con $z \in \mathbb{R}$, otteniamo $p_X(z) = \mathbf{P}(X \in A) = \mathbf{P}(Y \in A) = p_Y(z)$, dove la prima e la terza uguaglianza derivano dalla definizione di densità di probabilità , mentre la secondo deriva dall'ipotesi che X, Y hanno la stessa legge. Siccome per una generica v.a. discreta Z vale

$$\mathbf{P}(Z \in A) = \sum_{z \in A} p_Z(z)$$

con la convenzione che la somma a destra è per i valori z per cui $p_Z(z) > 0$, abbiamo che data una v.a. discreta la sua legge à determinata dalla sua densità di probabilità . Quindi se due v.a. discrete X, Y hanno la stessa densità di probabilità , allora devono avere la stessa legge.

La dimostrazione nel caso di v.a. continue richiede alcune conoscenze di teoria della misura per ora non acquisite dagli studenti. Ci limitiamo ad osservare che se f_X e f_Y sono uguali allora banalmente X, Y hanno la stessa legge. □

12. Interpretazione frequentistica della probabilità

Consideriamo una successione di prove identiche e operativamente indipendenti (es. lancio un dado infinite volte). Sia $(\hat{S}, \hat{P})$ lo spazio di probabilità che descrive una generica di queste prove (nell'esempio precendente: $\hat{S} = \{1,2,3,4,5,6\}$ e $\hat{P}(E) = |E|/6$ per ogni $E \subset \hat{S}$). Fisso un evento $F \subset \hat{S}$ (es. $F =$"esce 1 o esce 2").

Sia (S, P) lo spazio di probabilità che descrive la suddetta successione di prove. Considero le variabili aleatorie $X_n : S \to \{0,1\}$ con $n = 1,2,3,\ldots$ definite come

$$X_n := \begin{cases} 1 & \text{nella prova n–esima si verifica } F\,, \\ 0 & \text{altrimenti}\,. \end{cases}$$

Si noti che X_n assume valore 1 con prob. $\hat{P}(F)$ e assume valore 0 con prob. $\hat{P}(F^c)$. Quindi le variabili X_n, con $n \geq 1$, sono identicamente distribuite ed inoltre $E(X_n) =$

$\hat{P}(F)$. Siccome le variabili X_n, con $n \geq 1$, su riferiscono ad esperimenti operativamente indipendenti abbiamo che le variabili aleatorie X_n, con $n \geq 1$, sono indipendenti.

Notiamo che $\frac{X_1+X_2+\cdots X_n}{n}$= frequenza relativa con cui l'evento F si verifica nelle prime n prove.

Siccome abbiamo verificato che X_n, $n \geq 1$, sono variabili aleatorie i.i.d. con media $\hat{P}(F)$, possiamo applicare la legge forte dei grandi numeri che ci dà:

$$P\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots X_n}{n} = \hat{P}(F)\right) = 1\,.$$

Abbiamo quindi che con P–probabilità 1 il limite della frequenza relativa con cui l'evento F si verifica nelle prime n prove è dato da $\hat{P}(F)$. Questa affermazione è in accordo con l'interpretazione frequentistica della probabilità.

13. Complementi sulla funzione generatrice dei momenti

Ricordiamo il seguente fatto, di cui non diamo la dimostrazione (non sarà richiesta):

Proposizione 26. *Siano date due v.a. X, Y tali che in un opportuno intorno dell'origine valga $M_X(t) = M_Y(t) < \infty$. Allora X e Y hanno la stessa distribuzione.*

Usando il suddetto fatto si dimostra che la somma di due v.a. di Poisson indipendenti è v.a. di Poisson (di opportuno parametro) e analogamente la somma di due v.a. gaussiane indipendenti è v.a. gaussiana (di opportuni parametri) (vedasi es. 7g e 7h del Ross, sez. 7.7)