Calcolo delle Probabilità

Marco Casu

Contents

1 Il Modello Probabilistico

1.1 Il Problema dei Compleanni

Qaunte sono le probabilità che in un gruppo di 25 persone, 2 di queste siano nate lo stesso giorno? Procediamo nel **modellizzare** tale problema. Le possibili date di nascita sono 365.

$$PossibiliDate = \{1, 2, 3, 4..., 365\}$$

Intervistando un singolo individuo, questo può darmi 365 risposte, la probabilità che quindi un individuo di nome *Tizio* sia nato lo specifico giorno i è di 1 su 365.

$$\mathbb{P}(\textit{Tizio} \text{ è nato il giorno } i) = \frac{1}{365}$$

Tale probabilità si dice **uniforme** perchè tutti i 365 esiti hanno la stessa probabilità di $\frac{1}{365}$ di risultare. Adesso, quante sono le probabilità che *Tizio* sia nato il giorno i e *Caio* il giorno j?

$$\mathbb{P}(\textit{Tizio} \text{ è nato } i \text{ e } \textit{Caio} \text{ è nato } j) = \mathbb{P}(\textit{Tizio} \text{ è nato il giorno } i) \cdot \mathbb{P}(\textit{Caio} \text{ è nato il giorno } j)$$

Questo si chiama **concetto di indipendenza**, ed è ovvio che le probabilità sono precisamente $\frac{1}{365} \cdot \frac{1}{365} = \left(\frac{1}{365}\right)^2$. Quindie probabilità che 25 persone siano nate in 25 giorni pre determinati sono precisamente :

$$\mathbb{P}(a_1 \text{ è nato } i_1, a_2 \text{ è nato } i_2 ..., a_{25} \text{ è nato } i_{25}) = \left(\frac{1}{365}\right)^{25}$$

Queste sono le probabilità che 25 persone siano nate lo stesso giorno, a noi ci interessano però le probabilità che almeno 2 persone siano nate lo stesso giorno. Prima di fare ciò andiamo a definire quello che si dice **modello probabilistico**.
Chiamiamo Ω lo spazio degli eventi elementare (in seguito verrà definito in modo formale), Ossia l'insieme di tutti i possibili risultati dell'esperimento. Sia l'esperimento in questo caso, la data di nascita per k persone, il nostro Ω sarà l'insieme di tutte le possibili combinazioni lunghe 25 di numeri da 1 a 365, che sono precisamente $|\Omega| = 365^{25}$. La probabilità uniforme che un certo evento dello spazio degli eventi $\Omega = \omega_1, \omega_2 ..., \omega_k$ accada è $\mathbb{P}(\omega_i) = \frac{1}{|\Omega|}$. Definiamo adesso un evento **binario**, ossia un evento che ha come risposta, o si o no,ossia le probabilità che esso accada e che esso non accada. Definiamo l'evento A come l'evento in cui almeno 2 persone fanno il compleanno lo stesso giorno.

$$A = \{\omega \in \Omega : \exists \omega_i, \omega_j \in \omega | i \neq j \wedge \omega_i = \omega_j\}$$

Quindi tale A è un sotto-insieme di Ω in cui tutti gli elementi hanno almeno 2 date uguali. Presa con $|A|$ la cardinalità di tale sotto-insieme, le probabilità che tale evento accada sono :

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

Piuttosto che trovare l'evento in cui ci sono almeno 2 persone con lo stesso compleanno, è più facile trovare il suo **complementare**, ossia l'evento in cui nessuno fa il compleanno lo stesso giorno.

$$A^c = \{\omega \in \Omega | \forall i \neq j, \omega_i \neq \omega_j\}$$

Una volta trovata la probabilità di A^c, sapremo che $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$.

essendo $k = 25$ e $n = 365 \rightarrow |A^c| = n \cdot (n-1) \cdot (n-2) ... \cdot (n-k+1) \implies \mathbb{P}(A) = 1 - \frac{|A^c|}{365^{25}}$

Vediamo una tabella di possibili risultati al variare di k (numero di persone) :

k	4	16	22	23	63
$\mathbb{P}(A)$	0.06	0.28	0.42	0.50	0.99

Risulta chiaro come, intervistando 25 persone, c'è una probabilità superiore al 50 % che almeno due di esse condividano il compleanno.

1.2 Assiomi della probabilità

Definiamo adesso formalmente il **modello probabilistico**, esso ha 3 ingredienti :

1 - Lo Spazio degli Eventi Elementari
Detto anche *spazio campionario*, non è altro che l'insieme di tutti i possibili esiti dell' esperimento, e viene denominato con Ω. Ad esempio, se l'esperimento è il lancio di una moneta, si avrà $\Omega = \{\text{Testa,Croce}\}$. Nel caso dell'esperimento dei compleanni, lo spazio campionario sarà composto da tutte le possibili combinazioni lunghe k (numero di persone alla quale si chiede la data di nascita) di numeri da 1 a 365 (i giorni esistenti). Per la maggior parte dei casi analizzati in questo corso, Ω sarà un insieme finito. Si indica con $|\Omega|$ la cardinalità dello spazio campionario, ossia il numero dei suoi elementi.

2 - Algebra degli Eventi
Prima di definire l'algebra degli eventi, dichiariamo cos'è un **evento**, ossia una *domanda binaria* sull'esito dell'esperimento, in termini matematici, è un qualsiasi sotto-insieme Ω, compresi gli elementi singoli di Ω, detti **eventi elementari**. Se l'esperimento è il lancio di un dato, i possibili esiti sono $\Omega = \{1, 2, 3, 4, 5, 6\}$, vogliamo definire l'evento A come la probabilità che l'esito del lancio sia un numero maggiore di 4, esso non è altro che un sotto-insieme dello spazio campionario, ossia $A = \{\omega \in \Omega | \omega \geq 4\} = \{4, 5, 6\}$.
Su tali eventi sono definite delle operazioni insiemistiche di unione, intersezione e complemento.
L'Algebra degli Eventi, definita con il simbolo $\mathcal{A}$, è l'insieme chiuso[1] di *tutti i possibili sotto-insiemi* dello spazio campionario, ossia tutti i possibili eventi (anche non elementari) per la quale si può misurare la probabilità. *Ad esempio*, se prendiamo il caso del dado $\Omega = \{1, 2, 3, 4, 5, 6\}$, l'evento in cui il risultato sia 2 o 3, è l'unione dei due eventi elementari $A = \{2\}$ e $B = \{3\} \rightarrow C = A \cup B = \{2, 3\}$. Si noti che $|\Omega| = n \implies |\mathcal{A}| = 2^n$

3 - La Funzione Probabilità La probabilità di un determinato evento appartenente ad $\mathcal{A}$ è misurata tramite una funzione definita come $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$, essa associa ad ogni elemento dell'algebra degli eventi, una probabilità misurata tra 0 ed 1, dove 0 indica che l'evento è impossibile, ed 1 che l'evento è certo. La funzione $\mathbb{P}$ deve soddisfare 3 condizioni :

- $\mathbb{P}(\emptyset) = 0$ La probabilità dell'insieme vuoto è 0.
- $\mathbb{P}(\Omega) = 1$ La probabilità che si verifichi un evento di Ω è l'evento certo.
- $\mathbb{P}$ è una funzione *additiva*, ossia che per due eventi disgiunti A e B, nel senso che $A \cap B = \emptyset$, vale che $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Vediamo adesso **alcune considerazioni** riguardo gli assiomi appena enunciati.
Risulta ovvio che $\Omega = A \cup A^c \implies \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. Diversamente, presi due eventi A e B *non necessariamente disgiunti*, vale che $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Un'altra considerazione, è che $\mathbb{P}$ è una funzione *monotona* rispetto all'inclusione di insiemi, detto formalmente $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$.
Quando si vuole misurare la probabilità di un evento in un certo esperimento, bisogna quindi costruire un modello probabilistico, definendone gli eventi elementari e la probabilità ad essi associati. Possiamo quindi definire un altra funzione p, che ha come dominio esclusivamente

[1] È un insieme chiuso rispetto alle operazioni di unione, intersezione e complemento, ossia il risultato di tali operazioni su uno o più elementi dell'insieme, è sempre parte di tale insieme.

gli eventi elementari Ω, e non gli eventi di $\mathcal{A}$, essa sarà denominata come **probabilità degli eventi elementari**, $p : \Omega \to [0,1]$. Risulta chiaro che $\sum_{\omega \in \Omega} p(\omega) = 1$.

Ad esempio, facciamo un esperimento sul lancio di un dado truccato in cui è più probabile che esca il numero 6 piuttosto che altri :

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$p(1) = p(2) = p(3) = 0.1$
$p(4) = p(5) = 0.2$
$p(6) = 0.3$
Definiamo quindi $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$, e risulta chiaro che :

$$\sum_{\omega \in \Omega} p(\omega) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \mathbb{P}(\bigcup_{\omega \in \Omega} \{\omega\}) = 1$$

La probabilità in un modello probabilistico, si dice **uniforme** se $p(w) = \frac{1}{|\Omega|}$ $\forall \omega \in \Omega$, e significa che tutti gli eventi elementari hanno la stessa probabilità di accadere, da qui ne deriva che la probabilità di un evento $A \in \mathcal{A}$ è $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$.

2 Calcolo Combinatorio

Riassumiamo in questa sezione ciò che si è visto nel corso di *Metodi matematici per l'informatica*. Partiamo con un problema, si dia il caso che in un urna ci sono 3 palline bianche e 2 palline nere, qual'è la probabilità che, estraendo due palline, la seconda sia bianca? Formalizziamo il problema secondo il modello probabilistico. L'evento elementare è l'estrazione *ordinata* di due palline $\omega = (\omega_1, \omega_2)$, dove ω_1 = prima estratta e ω_2 = seconda estratta. Possiamo numerare le palline da 1 a 5, dicendo che le palline bianche son quelle da 1 a 3, e le restanti le nere. Detto ciò, possiamo capire che l'insieme degli eventi elementari, sono tutte le coppie di numeri naturali compresi fra 1 a 5 senza ripetizioni.

$$\Omega = \{(w_1, w_2) | w_1, w_2 \in \{1, 2, 3, 4, 5\}, w_1 \neq w_2\} \tag{1}$$

È di facile intuizione capire che $|\Omega| = 20$.
L'evento A invece rappresenta il sotto-insieme di Ω, dove 1,2 o 3 appaiano sempre come seconda coordinata dei suoi elementi, ossia tutte le estrazioni in cui la seconda pallina è bianca. Per calcolare $\mathbb{P}(A)$, dobbiamo capire qual'è prima la cardinalità di A, che risulta essere $|A| = 3 \cdot 4 = 12$, il 3 rappresenta le 3 possibili palline bianche della seconda estrazione, il 4 invece tutte le altre possibili rimanenti (per la quale non ci interessa il colore essendo la prima estrazione.) Fatto ciò si ha $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{12}{20} = \frac{3}{5} = 0.6$. La probabilità è quindi del 60%.

2.1 Principio fondamentale del calcolo combinatorio

Dati due insiemi finiti e non vuoti A e B, è possibile formare $|A| \cdot |B|$ diverse coppie ordinate prendendo un primo elemento da A ed un secondo elemento B.

$$|A \times B| = |A| \cdot |B| \tag{2}$$

Vediamo un *esempio* :
Lanciamo un dado 6 volte, quant'è la probabilità che esca almeno una delle volte 6? Diciamo che un evento elementare ω è una ennupla di 6 elementi (i lanci) composta da numeri da 1 a 6 (esiti del dado).

$$\Omega = \{(\omega_1, \omega_2, ...\omega_6)|\omega_i \in \{1,2,3,4,5,6\}\} \tag{3}$$

Risulta chiaro (per il principio moltiplicativo) che $|\Omega| = 6^6$. Dobbiamo adesso calcolare la cardinalità di $A = \{\omega \in \Omega | \exists \omega_i \in \omega \to \omega_i = 6\}$, ossia tutti gli eventi elementari in cui 6 compare almeno in un lancio. Ci risulta però più facile calcolare il suo complementare, ossia tutti gli eventi in cui 6 non compare mai, ossia ogni ennupla di 6 elementi (i lanci) composta da numeri da 1 a 5 (esiti del dado escluso il 6).

$$A^c = \{(\omega_1, \omega_2, ...\omega_6)|\omega_i \in \{1,2,3,4,5\}\} \tag{4}$$

Risulta chiaro (per il principio moltiplicativo) che $|A^c| = 5^6$. Avendo la $|A^c|$, possiamo trovare la probabilità dell'evento A :

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{|A^c|}{|\Omega|} = 1 - \frac{5^6}{6^6} \simeq 0.66 \tag{5}$$

2.2 Modelli del calcolo combinatorio

Prima di parlare dei modelli del calcolo combinatorio, introduciamo un concetto importante :

Permutazione
Sia S un insieme finito e non vuoto, una *permutazione* di S è una scelta di ordine tra gli elementi di S. Le possibili permutazioni risultano essere $|S|!$.

Passiamo adesso alla presentazione di alcuni modelli fondamentali nella quale spesso ricadono i problemi di conteggio. I problemi verranno *metaforizzati* secondo il problema delle palline nelle urne.

2.2.1 Caso 1 - Estrazioni Ordinate con Rimpiazzo

Si hanno n palline dentro un urna e si vogliono fare k estrazioni con rimpiazzo, ossia, una volta pescata una pallina, si reinserisce nell'urna (un elemento stesso può essere estratto più volte). Si tiene conto dell'ordine, quindi $\{1,2,3\} \neq \{1,3,2\}$. Per problemi di questo tipo, vale che :

$$\Omega = \{(\omega_1, \omega_2 ..., \omega_k)|\omega_i \in \{1,2,3...,n\}\} \implies |\Omega| = n^k \tag{6}$$

2.2.2 Caso 2 - Estrazioni Ordinate senza Rimpiazzo

Si hanno n palline dentro un urna e si vogliono fare k estrazioni senza rimpiazzo, ossia, una volta pescata una pallina, essa non potrà essere ripescata (gli elementi di ogni estrazione sono distinti). Si tiene conto dell'ordine, quindi $\{1,2,3\} \neq \{1,3,2\}$. Per problemi di questo tipo, vale che :

$$\Omega = \{(\omega_1, \omega_2 ..., \omega_k)|\omega_i \in \{1,2,3...,n\} | i \neq j \implies w_i \neq w_j\} \implies |\Omega| = \frac{n!}{(n-k)!} \tag{7}$$

Esempio :
7 persone si trovano in un ascensore dentro un palazzo di 10 piani e si apprestano a scendere su un piano, quante sono le probabilità per cui tutte quante scendano su piani diversi?
L'Evento elementare è composta da una ennupla di 7 elementi (le persone) composta da numeri che vanno da 1 a 10 (il piano in cui scendono). Quindi :

$$\Omega = \{(\omega_1, \omega_2..., \omega_7)|\omega_i = \{1,2,3,4,5,6,7,8,9,10\}\}$$

Dati i modelli precedenti risulta chiaro che $|\Omega| = 10^7$ e $A = \{$Ognuno scende su un piano diverso$\}$ ha cardinalità $|A| = \dfrac{10!}{3!} \implies \mathbb{P}(A) = \dfrac{\frac{10!}{3!}}{10^7} \simeq 0.06$.

2.2.3 Caso 3 - Estrazioni non Ordinate senza Rimpiazzo

Si hanno n palline dentro un urna, se ne estraggono k, ma senza tener conto dell'ordine delle palline, ad esempio, se ho le palline denominate 1,2 e 3, estraendole, si ha che le estrazioni 123, 321 ,132, 312, 231, 213 sono lo stesso evento. Come rappresento tale spazio elementare Ω a livello insiemistico? Si consideri lo spazio delle *Estrazioni Ordinate senza Rimpiazzo*, ossia:

$$A = \{(\omega_1, \omega_2..., \omega_k)|\omega_i \in \{1,2,3...,n\}|i \neq j \implies w_i \neq w_j\}$$

Stabiliamo una relazione di equivalenza su A, denominata con il simbolo $\sim$, per cui vale che $\omega \sim \omega'$ se e solo se differiscono solo per l'ordine.
Esempio : $\omega = \{1,2,3\}, \omega' = \{3,1,2\} \iff \omega \sim \omega'$. L'insieme degli eventi elementari per rappresentare le *Estrazioni non Ordinate senza Rimpiazzo* non è altro che l'insieme quoziente di A su $\sim$, ossia l'insieme di tutte le sue classi di equivalenza.

$$\Omega = {}^{A}/_{\sim}$$

Possiamo scegliere come rappresentanti delle classi di equivalenza, quegli elementi in cui le palline sono ordinate in ordine crescente, in tal modo, lo spazio elementare è anche rappresentabile come :

$$\{(\omega_1, \omega_2..., \omega_k)|\omega_i \in \{1,2,3...,n\} \wedge i < j \implies w_i < w_j\}$$

Qual'è però la cardinalità di Ω in questo caso? Possiamo considerare la cardinalità di A, divisa per il numero di elementi in ogni classe di equivalenza. Essendo che ogni classe di equivalenza non è altro permutazione di k elementi, esse avranno ognuna $k!$ elementi, da qui risulta chiara la cardinalità del nostro spazio degli eventi :

$$\text{sia } [\omega] \in {}^{A}/_{\sim}, \text{ si ha } |\Omega| = \frac{|A|}{|[\omega]|} = \frac{n!}{k! \cdot (n-k)!} \text{ si denota come } \binom{n}{k} \tag{8}$$

Tradotto in linguaggio umano, $\binom{n}{k}$ indica il numero di modi di scegliere k elementi da un insieme n. Tale conteggio è anche detto **combinazione semplice**. Vale la proprietà :

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$$

Binomio di Newton

$$\text{Siano } a, b \in \mathbb{R} \text{ e } n \in \mathbb{N}, \text{ vale che } (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot a^{n-k} \cdot b$$

2.2.4 Caso 4 - Estrazioni non Ordinate con Rimpiazzo

Consideriamo il lancio di 2 dadi identici, senza considerare l'ordine. Abbiamo che l'evento elementare è la coppia di numeri da 1 a 6 (senza contare l'ordine, quindi consideriamo la classe di equivalenza in cui i due dadi hanno esito in ordine crescente), lo spazio risulta quindi : $\Omega = \{(\omega_1, \omega_2) | 1 \leq \omega_1 \leq \omega_2 \leq 6\}$
Se i due dadi fossero distinti e l'ordine avesse importanza, i possibili eventi sarebbero 6^2, ma in questo caso, non risulta la scelta naturale, dato che gli eventi (a, b) e (b, a) sono identici.

È il caso di un *estrazione non ordinata con rimpiazzo*, dato che estraggo due palline, senza contare l'ordine, con la possibilità di poter estrarre due volte la stessa pallina.
Se ho n palline e faccio k estrazioni, il numero di estrazioni possibili non ordinate con rimpiazzo risulta essere :

$$\binom{n+k-1}{k} \tag{9}$$

Tornando al problema dei dadi, i possibili esiti sono :

$$\binom{6+2-1}{2} = \frac{(6+2-1)!}{2!(6+2-1-2)!} = \frac{7!}{2!5!} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 2} = \frac{7 \cdot 6}{2} = 21$$

2.3 Principio di Esclusione/Inclusione

Siano $A, B, C \subset \Omega$ tre eventi non necessariamente disgiunti di uno spazio di probabilità, se volessi calcolare la probabilità di $A \cup B$, dovrei fare la somma delle probabilità di A e B, sottraendo la probabilità dell'intersezione. Già con 3 eventi, se volessi la probabilità dovrei fare :

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - (\mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C)) + \mathbb{P}(A \cap B \cap C)$$

Quando si parla quindi di un numero elevato di eventi, il calcolo risulta difficile, ed è descritto dalla seguente formula :

$$\mathbb{P}(A_1 \cup A_2 ..., \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \cdot \sum_{1 < i_1 < i_2 ..., < i_k < n} \mathbb{P}(A_{i_1} \cap A_{i_2} ..., \cap A_{i_k}) \tag{10}$$

2.4 Il Multinomio

Esaminiamo il problema in cui si vogliono distribuire n oggetti distinti in k scatole distinte, in modo che ciascuna di esse contenga nell'ordine $n_1, n_2 ..., n_k$ oggetti, per cui $\sum_{i=1}^{n} n_i = n$. In quanti modi posso effettuare tale suddivisione? Ci sono $\binom{n}{n_1}$ scelte per la prima scatole, $\binom{n-n_1}{n_2}$ per la seconda e così via, fino a $\frac{n!}{n_1! \cdot n_2! ..., \cdot n_k!}$, che si denota con $\binom{n}{n_1, n_2 ..., n_k}$, e rappresenta il numero di suddivisioni di n oggetti distinti in k gruppi distinti con, rispettivamente $n_1, n_2 ..., n_k$ oggetti.

Esempio : Vi sono 10 poliziotti, se 5 di essi devono pattugliare due strade, 2 di essi restare in stazione ed altri 3 stare in riserva, quanti sono i modi possibili di assegnare i 3 compiti ai 10 poliziotti?

$$\binom{10}{5, 2, 3} = \frac{10!}{5! \cdot 2! \cdot 3!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 3 \cdot 2 \cdot 2} = \frac{30240}{12} = 2520$$

2.5 Problema degli Accoppiamenti

Poniamo un esempio, c'è una festa con n invitati, ogni invitato porta con se un ombrello, al termine della festa ogni invitato se ne va prendendo un ombrello a caso, qual'è la probabilità che nessun invitato riprenda il proprio ombrello? Se il numero di persone n tende ad infinito, come si comporta tale probabilità?

Si dice che in uno spazio di probabilità, se un evento elementare composto da una lista $a_1, a_2..., a_n$ che possono assumere valori da 1 ad n, se un elemento a_i assume valore i, esso si dice **punto fisso**, quindi possiamo tradurre il problema degli ombrelli (considerando ogni persona come un elemento dell'evento elementare, ed i valori che possono assumere come gli ombrelli) come : Qual'è la probabilità che una permutazione di n non abbia punti fissi?

$$\Omega = \{(\omega_1, \omega_2..., \omega_n)|i \neq j \implies \omega_i \neq \omega_j|\omega_k \in \{1, 2, 3..., n\}\} \text{ dove } |\Omega| = n!$$
$$B = \{\text{permutazioni senza punti fissi}\} = \{\omega \in \Omega|\forall i \in \{1, 2, 3..., n\}\omega_i \neq i\}$$

Poniamo il complementare di B, ossia $B^c = A = \{\text{la permutazione ha almeno un punto fisso}\}$, e distinguiamolo in n sotto eventi, ossia $A_1 = \{\omega_1 \text{ è un punto fisso}\}$ poi $A_2 = \{\omega_2 \text{ è un punto fisso}\}$ e così via per ogni punto. Sapremo così che $A = A_1 \cup A_2..., \cup A_n$. Calcoliamo la cardinalità di un singolo evento A_i, risulta facile, dato che poniamo un punto fisso su $\omega_i = i$ e poi il resto è una permutazione normale. $|A_i| = (n-1)!$, quindi $\mathbb{P}(A_i) = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$.
Quindi per calcolare la probabilità di A dobbiamo sommare le probabilità di ogni A_i, ma ovviamente tali sotto-eventi *non sono disgiunti*, quindi dobbiamo sottrarre le intersezioni. Notiamo come $|A_i \cap A_j| = (n-2)!$, perchè fissiamo entrambi i punti i, j e ed il resto è una semplice permutazione. Questo fino a k intersezioni : $|A_1 \cap A_2..., \cap A_k| = (n-k)!$. Quindi : $\mathbb{P}(A_1 \cap A_2..., \cap A_k) = \dfrac{(n-k)!}{n!}$. A questo punto applichiamo qui il **principio di esclusione/inclusione** :

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = \sum_{i=1}^{n}(-1)^{k+1} \sum_{1<i_1<i_2...,<i_k<n} \mathbb{P}(A_1 \cap A_2..., \cap A_k) \tag{11}$$

Ma sappiamo che $\mathbb{P}(A_1 \cap A_2..., \cap A_k) = \dfrac{(n-k)!}{n!}$ quindi riscriviamo :

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = \sum_{i=1}^{n}(-1)^{k+1} \sum_{1<i_1<i_2...,<i_k<n} \frac{(n-k)!}{n!} \tag{12}$$

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = \sum_{i=1}^{n}(-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} \tag{13}$$

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = \sum_{i=1}^{n}(-1)^{k+1} \frac{n!}{(n-k)!k!} \frac{(n-k)!}{n!} \tag{14}$$

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = \sum_{i=1}^{n}(-1)^{k+1} \frac{1}{k!} \tag{15}$$

Essendo A il complementare di B :

$$\mathbb{P}(B) = 1 - \sum_{i=1}^{n}(-1)^{k+1} \frac{1}{k!} \tag{16}$$

Con n che tende ad infinito si ha :

$$\lim_{n\to\infty} \mathbb{P}(B) = 1 - \sum_{i=1}^{n} (-1)^k \frac{1}{k!} = \frac{1}{e} \tag{17}$$

Quindi la probabilità che in una permutazione non vi siano punti fissi è **costante**, e vale esattamente $\frac{1}{e}$.

3 Spazi di Probabilità Notevoli

3.1 Spazi di Probabilità Prodotto

Consideriamo due esperimenti :

- Lancio una moneta equa
- Viene misurata la temperatura in Brasile (che può variari dai 10 ai 50 gradi)

Per i due esperimenti diversi abbiamo due spazi di probabilità differenti, $(\Omega_1, \mathbb{P}_1)$ e $(\Omega_2, \mathbb{P}_2)$, considerando l'esperimento congiunti, essendo essi totalmente indipendenti (non si influenzano fra loro), possiamo immaginare ogni evento come una coppia composta dal primo evento elementare ed il secondo evento elementare, costruendo un nuovo spazio elementare come prodotto cartesiani dei primi due : $\Omega = \Omega_1 \times \Omega_2$.

$$\Omega = \{(\omega_1, \omega_2) | \omega_1 \in \{T, C\}, \omega_2 \in \mathbb{N} | 10 \leq \omega_2 \leq 50$$

Se i due esperimenti non si influenzano, la scelta naturale è la **probabilità prodotto** :

$$\mathbb{P}(\{\omega\}) = \mathbb{P}(\{(\omega_1, \omega_2)\}) = \mathbb{P}_1(\{\omega_1\}) \cdot \mathbb{P}_2(\{\omega_2\}) \tag{18}$$

Devo verificare che $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$, ho che :

$$\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} \mathbb{P}_1(\{\omega_1\}) \cdot \mathbb{P}_2(\{\omega_2\}) \tag{19}$$

$$\sum_{\omega_1 \in \Omega_1} \mathbb{P}_1(\{\omega_1\}) \cdot \sum_{\omega_2 \in \Omega_2} \mathbb{P}_2(\{\omega_2\}) = 1 \cdot 1 = 1 \quad \blacksquare \tag{20}$$

Se $\mathbb{P}_1$ e $\mathbb{P}_2$ sono probabilità uniformi su Ω_1 e Ω_2, allora $\mathbb{P} = \mathbb{P}_1 \cdot \mathbb{P}_2$ è uniforme su $\Omega = \Omega_1 \cdot \Omega_2$. Si dimostra facilmente :

$$\mathbb{P}(\{\omega_1, \omega_2\}) = \mathbb{P}_1(\omega_1) \cdot \mathbb{P}_2(\omega_2) = \frac{1}{|\Omega_1|} \cdot \frac{1}{|\Omega_2|} = \frac{1}{|\Omega_1 \times \Omega_2|} = \frac{1}{|\Omega|}$$

3.1.1 Indipendenza di Eventi

Sia $\mathbb{P}$ la funzione probabilità, che in un certo modello probabilistico soddisfa la seguente condizione di *compatibilità* : Sia $A \subset \Omega$ un evento particolare, per cui posso decidere l'esito dell'esperimento osservando esclusivamente $(\Omega_1, \mathbb{P}_1)$, tale che $\forall A_1 \in \Omega_1$, $A = A_1 \times \Omega_2$. In matematica, tale A è detto *cilindro*. Vale che $\mathbb{P}(A) = \mathbb{P}_1(A)_1$.

$$\mathbb{P}(A) = \sum_{(\omega_1,\omega_2)\in A_1\times\Omega_2} \mathbb{P}_1(\omega_1)\cdot\mathbb{P}_2(\omega_2) = \sum_{\omega_1\in A_1}\mathbb{P}_1(\omega_1)\cdot\sum_{\omega_2\in\Omega_2}\mathbb{P}_2(\omega_2) = \sum_{\omega_1\in A_1}\mathbb{P}_1(\omega_1)\cdot 1 = \mathbb{P}_1(A)_1$$

Se esiste un evento B, che è un evento particolare, per cui posso decidere l'esito dell'esperimento osservando esclusivamente $(\Omega_2, \mathbb{P}_2)$ (analogamente al primo caso), risulta che $\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$. Tali eventi non si *influenzano* fra loro :

> In un modello probabilistico $(\Omega, \mathbb{P})$, due eventi $A, B \subseteq \Omega$ sono detti **indipendenti** se :
>
> $$\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$$

Tale relazione di *indipendenza* è data esclusivamente dalla funzione probabilità, non è una proprietà insiemistica.

Esempio : Estraggo a caso una carta fra un mazzo di 40. $\Omega = \{\omega | 1 \leq \omega \leq 40\}$, ed ho i due eventi $A = \{$ Estraggo un 7 $\}$ e
$B = \{$ Estraggo una carta di denara $\}$, tali eventi sono indipendenti, dato che l'estrazione di un 7, in nessun modo condiziona la probabilità di estrarre una carta di denara, e viceversa. Infatti $\mathbb{P}(A)\cdot\mathbb{P}(B) = \frac{1}{10}\cdot\frac{1}{4} = \frac{1}{40} = \mathbb{P}(A\cap B)$.

Discutiamo adesso **l'indipendenza fra 3 eventi**. $A, B, C \subseteq \Omega$, si dicono *indipendenti* se sono soddisfatte le seguenti condizioni :

- $\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B), \mathbb{P}(A\cap C) = \mathbb{P}(A)\cdot\mathbb{P}(C), \mathbb{P}(C\cap B) = \mathbb{P}(C)\cdot\mathbb{P}(B)$
- $\mathbb{P}(A\cap B\cap C) = \mathbb{P}(A)\cdot\mathbb{P}(B)\cdot\mathbb{P}(C)$

Passiamo al caso generale, n eventi $A_1, A_2, A_3 ..., A_n$ si dicono **indipendenti** quando, presa una qualunque sotto-famiglia di tali insiemi, la probabilità di tale intersezione *fattorizza* :

$$\forall k | 1 \leq i_1 < i_2 ..., < i_k \leq n,\ \mathbb{P}(A_{i_1}\cap A_{i_2} ..., \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) ..., \cdot\mathbb{P}(A_{i_k})$$

Ad esempio, in uno schema di probabilità prodotto, se prendo $A_1 \subset \Omega_1, A_2 \subset \Omega_2 ..., A_n \subset \Omega_n$, risulta che tali eventi sono tutti indipendenti fra loro.

3.2 Schema di Bernoulli

Apriamo una digressione su un noto schema probabilistico. Vi è un *esperimento binario*[2] ripetuto n volte, gli esiti di tale risultato vengono codificati con 0 ed 1, quindi si ha :
$\Omega = \{(\omega_1, \omega_2 ... \omega_n) | \omega_i \in \{0,1\}\}$. Lo schema è quello di uno spazio di probabilità prodotto, dove $(\Omega_i, \mathbb{P}_i)$ definisce ogni lancio . Si definisce una costante reale $p \in [0,1]$, tale che, per ogni singolo lancio risulta che :

- $\mathbb{P}_i(\{1\}) = p$
- $\mathbb{P}_i(\{0\}) = 1 - p$

Ovviamente se $p = \frac{1}{2}$, la probabilità è uniforme. Scegliendo diversi lanci, risulta che :

[2] Un esperimento con due esiti (ad esempio il lancio di una moneta)

$$n = 2 \implies \Omega = \{00, 01, 10, 11\} \text{ con } \mathbb{P}(00) = (1-p)^2,\ \mathbb{P}(11) = p^2,\ \mathbb{P}(01) = \mathbb{P}(10) = p(1-p)$$

$$n = 3 \implies \Omega = \{000, 001, 010, 011, 100, 101, 110, 111\} \text{ con } \mathbb{P}(000) = (1-p)^3,\ \mathbb{P}(111) = p^3 \text{...ecc}$$

In generale, per n lanci risulta che :

$$\text{Sia } A \in \Omega : \mathbb{P}(A) = p^{\text{occorrenze di 1 in } A} \cdot (1-p)^{\text{occorrenze di 0 in } A}$$

Qui risulta comoda la codifica binaria perchè posso scrivere tale probabilità come :

$$\mathbb{P}(\{(\omega_1, \omega_2 ... \omega_n)\}) = p^{\displaystyle\sum_{i=1}^{n} \omega_i}$$

Per un evento A in cui si ottiene k volte 1 ed $n-k$ volte 0, la sua probabilità risulta essere :

$$\mathbb{P}(A) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

Tale formula è detta **distribuzione binomiale**.

3.3 Probabilità Condizionata

Vie è una scatola con dentro 3 palline bianche e 2 nere, si fanno 2 estrazioni ordinate senza rimpiazzo. Qual'è la probabilità che, dopo averne pescata una bianca, se ne peschi un'altra bianca? Vogliamo sapere la probabilità di un evento A, dando però una condizione, ossia che si sia già causato l'evento B.

> **Definizione Importante** : Consideriamo uno spazio di probabilità $\Omega, \mathbb{P}$, e siano $A, B \subseteq \Omega$, denominiamo $\mathbb{P}(A|B)$ la probabilità che A si verifichi, con la condizione che B si sia già verificato, e vale esattamente $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Osservazione : Se la probabilità è uniforme, si ha che $\mathbb{P}(A|B) = \dfrac{|A \cap B|}{|B|}$.

Esempio : Si dia il caso in cui si lancia 2 volte una moneta, qual'è la probabilità che al secondo lancio esca testa, sapendo che al primo lancio è uscito testa? $E = \{TT\}$ è l'evento in cui esce entrambe volte testa, ed $F = \{TT, TC\}$ l'evento corrispondente all'uscita di testa nel primo lancio. Si ha che $\mathbb{P}(E|F) = \dfrac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \dfrac{1/4}{2/4} = \dfrac{1}{2}$.

3.3.1 Formula di Probabilità Composte

Possiamo usare il concetto di probabilità condizionata per esprimere la probabilità di un evento, a seguito di una sequenza di eventi che si condizionano fra loro, si dia il caso che in una scatola vi sono b palline bianche e n palline nere, faccio k estrazioni. Chiamiamo $A_1 =$\{la prima è nera\}, $A_2 =$\{la seconda è nera\} fino ad $A_k =$\{la k-esima è nera\}. La probabilità che tutte le palline siano nere, sarà uguale alla composizione di tutti questi eventi riguardo la $i-$esima estrazione, ma si noti come, l'evento in cui le prime i palline sono nere, è uguale alla probabilità che la $i-$esiama sia nera con la condizione che le $i-1$ palline precedenti siano nere.

$$\mathbb{P}(\{\text{tutte nere}\}) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) ... \cdot \mathbb{P}(A_k|A_1 \cap ..., A_{k-1}) = \tag{21}$$

$$= \mathbb{P}(A_1)\frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \cdot ... \cdot \frac{\mathbb{P}(A_1 \cap A_2 \cap ..., A_k)}{\mathbb{P}(A_1 \cap A_2 \cap ..., A_{k-1})} = \text{ per semplificazione } = \tag{22}$$

$$= \mathbb{P}(A_1 \cap A_2 \cap ..., A_k) = \mathbb{P}(\bigcap_{i=1}^{k} A_i) \tag{23}$$

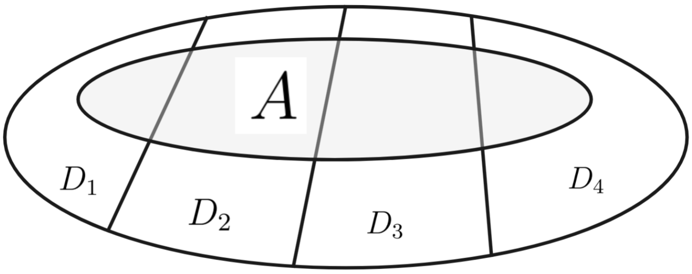
3.3.2 Formula di Probabilità Totale

Vediamo adesso il caso in cui, voglio che si verifichi un determinato evento, che può avere probabilità diverse a seconda della condizione verificatasi prima.
Vediamo un *esempio* : Si lancia una moneta equa, se esce testa, si pesca una pallina da un sacco contenente 2 palline nere e 3 bianche, se esce croce, si pesca una pallina da un sacco contenente 2 palline nere ed una bianca. È chiaro che, l'esito della moneta, condiziona la probabilità che la pallina estratta sia bianca. Se le condizioni che possono influire sulla probabilità dell'evento sono *disgiunte*, in questo caso lo sono perchè, o esce testa, o esce croce, la probabilità sarà uguale a :

$$\mathbb{P}(\{\text{si estrae una bianca}\}) = \mathbb{P}(T)\mathbb{P}(\{\text{esce bianca}\}|T) + \mathbb{P}(C\mathbb{P}(\{\text{esce bianca}\}|C)) \tag{24}$$

In generale, sia $\Omega, \mathbb{P}$ uno spazio di probabilità, se $D_1, D_2, D_3..., D_k$ è una partizione di Ω, tale che $\bigcup_{i=1}^{k} D_i = \Omega$, e sia $A \subseteq \Omega$ un evento che segue uno degli eventi D_i, la probabilità di A vale :

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(D_i)\mathbb{P}(A|D_i) \tag{25}$$



3.3.3 Formula di Bayes

Questo è il caso analogamente opposto a quello visto precedentemente, sia $D_1, D_2, D_3..., D_k$ una partizione di Ω, e sia $A \subseteq \Omega$ un evento che segue uno degli eventi D_i, sapendo che si è verificato A, possiamo calcolare la probabilità che si sia verificato un singolo D_i :

$$\mathbb{P}(D_i|A) = \frac{\mathbb{P}(D_i) \cdot \mathbb{P}(A|D_i)}{\sum_{j=1}^{k-1} \mathbb{P}(D_j) \cdot \mathbb{P}(A|D_j)} \tag{26}$$

Osservazione : Se A, B sono eventi disgiutni, allora $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Vediamo adesso un esempio facile di applicazione della formula delle probabilità totali, per ricavare il valore di una probabilità appartenente non banale. Si consideri lo schema di Bernoulli 3.2, su n lanci, qual'è la probabilità che esca testa, un numero pari di volte? Voglio procedere *ricorsivamente* in n, similmente a come si risolvono le equazioni di ricorrenza viste nel corso di Algoritmi 1.
Definisco $A_n = \{\text{Su } n \text{ lanci, testa esce un numero pari di volte}\}$ e $p_n = \mathbb{P}(A_n)$. Procedo utilizzando la formula delle probabilità totali, definendo le partizioni :

- $D_1 = \{\omega \in \Omega | \omega_1 = 0\}$ - al primo lancio esce croce, ovviamente $\mathbb{P}(D_1) = 1 - p$
- $D_2 = \{\omega \in \Omega | \omega_1 = 1\}$ - al primo lancio esce testa, ovviamente $\mathbb{P}(D_2) = p$

Ed ottengo :

$$p_n = \mathbb{P}(D_1)\mathbb{P}(A_n|D_1) + \mathbb{P}(D_2)\mathbb{P}(A_n|D_2) \tag{27}$$

Dove $\mathbb{P}(A_n|D_1)$ rappresenta la probabilità che Su $n-1$ lanci, testa esce un numero pari di volte, e $\mathbb{P}(A_n|D_2)$ rappresenta la probabilità che Su $n-1$ lanci, testa esce un numero dispari di volte. Quindi l'equazione si può riscrivere come :

$$p_n = (1-p)p_{n-1} + p(1-p_{n-1}) \tag{28}$$

Abbiamo definito una relazione *ricorsiva*, che tramite passaggi algebrici può essere espressa in forma esplicita.

3.3.4 Mappa sulla Probabilità Condizionata

Quando si vuole calcolare la probabilità di un evento con la condizioni che un'altro evento si sia verificato, è possibili applicare la formula che abbiamo visto all'inizio di questo paragrafo. Un altra via da seguire però, può essere quella di considerare un **nuovo spazio di probabilità**, in cui si considerano come eventi elementari, esclusivamente quelli che già includono che la condizione si sia verificata.

Se A è l'evento condizionante, posso considerare : l'applicazione:

$$B \subset \Omega, \ B \to \mathbb{P}(B|A)$$

Ossia la mappa che ad ogni evento di Ω, associa la sua probabilità con condizione A verificata. Tale mappa, è ancora una funzione di probabilità, sullo spazio degli eventi A, ciò è di facile dimostrazione:

$$\forall B \subset A, \ \mathbb{P}(B|A) \in [0,1] \tag{29}$$

$$\begin{cases} \mathbb{P}(\emptyset|A) = 0 \\ \mathbb{P}(A|A) = \dfrac{\mathbb{P}(A \cap A)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1 \end{cases} \tag{30}$$

Inoltre, tale funzione è additiva :

$$B_1, B_2 \subset A, \ \mathbb{P}(B_1 \cup B_2|A) = \mathbb{P}(B_1|A) + \mathbb{P}(B_2|A) \tag{31}$$

$$\frac{\mathbb{P}((B_1 \cup B_2) \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}((B_1 \cap A) \cup (B_2 \cap A))}{\mathbb{P}(A)} \implies \text{per additività di } \mathbb{P} \implies$$
$$\frac{\mathbb{P}(B_1 \cap A)}{\mathbb{P}(A)} + \frac{\mathbb{P}(B_2 \cap A)}{\mathbb{P}(A)}$$

Osservazione 1 : Se $\mathbb{P}$ è la probabilità uniforme su Ω, preso l'evento fissato $A \subset \Omega$, l'applicazione $B \subset \Omega, \ B \to \mathbb{P}(B|A)$ è anche essa ancora una probabilità uniforme.

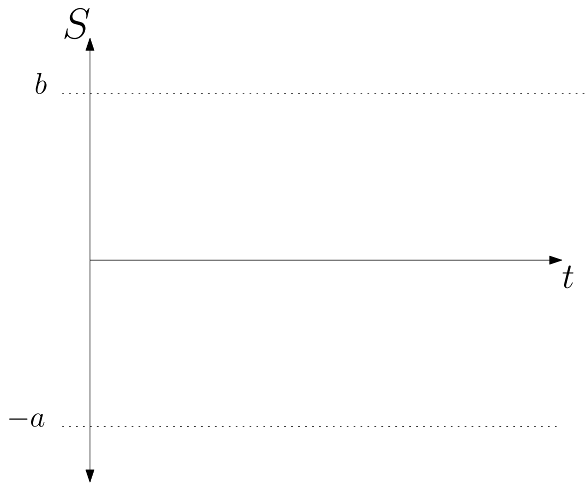
3.4 Passeggiata Aleatoria

3.4.1 La Rovina del Giocatore

Introduciamo adesso un'altro problema meno banale rispetto a quello appena visto, che richiede sempre l'applicazione della formula delle probabilità totali. Ci sono due giocatori A e B, che giocano ad un gioco generico, scomemttendo dei soldi. Ogni volta che un giocatore vince, riceve 1 euro dall'altro. A ha probabilità di vittoria uguale a p, e B ha probabilità di vittoria uguale a $1-p$, A inizia il gioco con un capitali pari ad a euro, e B inizia con un capitale pari a b euro. Giocano senza sosta, finche uno dei due giocatori non si rovina (finisce tutto il capitale). Qual'è la probabilità che A oppure B si rovini?

3.4.2 Modellizzazione

Tale problema ha un interpretazione geometrica ben precisa, e viene modellizzato in uno schema definito come *passeggiata aleatoria*. Si consideri il piano $\mathbb{Z}^+ \times \mathbb{Z}$:

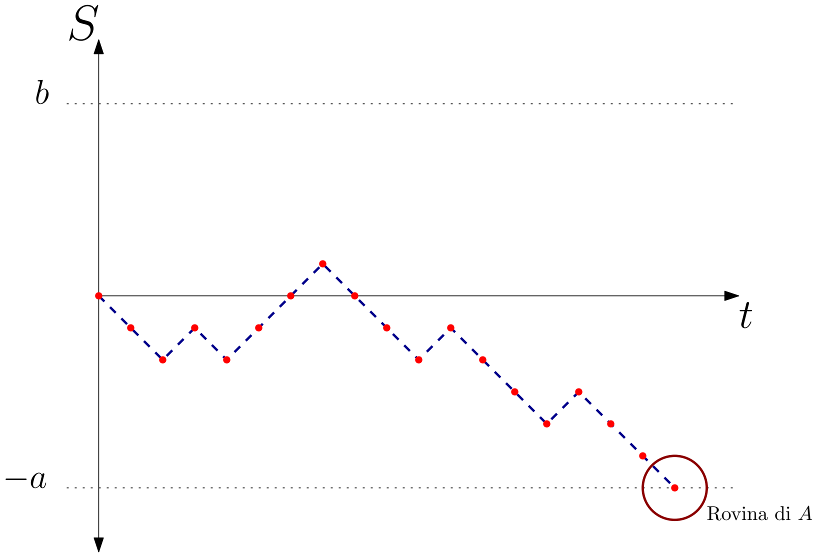


l'asse delle ascisse t rappresenta il tempo, o il numero di giocate fatte, mentre quello delle ordinate S rappresenta la somma in denaro vinta da A, ci sono poi due asintoti orizzontali nelle posizioni $S = b$ e $S = -a$. Ad ogni coppia t, S è associato un valore denominato S_n, che identifica l'istante di gioco, ossia la somma vinta (o persa se in negativo) da A al tempo $t = n$. All'avanzare del tempo, ossia delle giocate, la crescita o discesa di S_n dipende dalle probabilità che A vinca o no, se A vince al tempo i, nel tempo $i + 1$, S_n crescerà di un'unità, altrimenti decrescerà.

- $t = 0 \implies S_0 = 0$
- $t = 1 \implies S_1 = \begin{cases} +1 \text{ con probabilità } p \\ -1 \text{ con probabilità } 1-p \end{cases}$
- $t = 2 \implies S_2 = S_1 + \begin{cases} +1 \text{ con probabilità } p \\ -1 \text{ con probabilità } 1-p \end{cases}$
- ...
- $t = n+1 \implies S_{n+1} = \sum_{i=1}^{n} S_i + \begin{cases} +1 \text{ con probabilità } p \\ -1 \text{ con probabilità } 1-p \end{cases}$

Quindi nel grafico, S_n "spostandosi" verso destra, oscillerà avvicinandosi sempre di più o a b o a $-a$, con ovvio significato :

- $S_n = b \implies$ il giocatore B si è rovinato, ha vinto A.
- $S_n = -a \implies$ il giocatore A si è rovinato, ha vinto B.



La probabilità che A si rovini è quindi equivalente alla probabilità che S_n raggiunga $-a$, si può calcolare tale valore utilizzando la formula delle probabilità totali.

Introduciamo un nuovo parametro, ossia x, che equivale ad S_0, ossia quanto vale S all'inizio del gioco, fin'ora, abbiamo dato per scontato che si partisse sempre da una condizione di parità, dove $x = 0$. Definiamo quindi $\alpha(x)$ la probabilità che S_n raggiunga $-a$ da un punto di partenza x, e $\beta(x)$ la probabilità che S_n raggiunga b da un punto di partenza x. Ovviamente se $x = 0$ siamo nel caso iniziale, ma ad ogni passo, x incrementerà o decrementerà di 1. Definiamo quindi $S_n^{(x)}$ come la somma vinta da A nell'istante $t = n$, dove però $S_0^{(x)} \neq 0 = x$. Vediamo come si definisce in questo caso una relazione ricorsiva, utilizzo la formula della probabilità totale :

$$\beta(x) = \mathbb{P}(S_n+=1) \cdot \mathbb{P}(B \text{ si rovina } |S_n+=1) + \mathbb{P}(S_n-=1) \cdot \mathbb{P}(B \text{ si rovina } |S_n-=1) \tag{32}$$

Ne consegue :

$$\beta(x) = p \cdot \beta(x+1) + (1-p) \cdot \beta(x-1) \text{ finché } x = -a \vee x = b \tag{33}$$

Abbiamo quindi definito una relazione ricorsiva :

$$\begin{cases} \beta(x) = p \cdot \beta(x+1) + (1-p) \cdot \beta(x-1) \\ \beta(b) = 1 \\ \beta(-a) = 0 \end{cases}$$

Sviluppiamo il calcolo per ottenere la formula esplicita, iniziando a calcolare l'incremento :

$$(p+(1-p))\beta(x) = p \cdot \beta(x+1) + (1-p) \cdot \beta(x-1) \iff p(\beta(x+1) - \beta(x)) = (1-p)(\beta(x) - \beta(x-1)) \tag{34}$$

$$\iff \beta(x+1)-\beta(x) = \frac{1-p}{p}(\beta(x)-\beta(x-1)) \tag{35}$$

Gli incrementi sono quindi *proporzionali* e non lineari (simile alla funzione esponenziale). Avendo trovato adesso il valore ad ogni incremento generico, procediamo calcolando l'incremento ad ogni passo :

- passo 0) $\beta(-a) = 0$
- passo 1) $\beta(-a+1) = \beta(-a+1) - \beta(-a) = \beta(-a+1) - 0 = \gamma$

Abbiamo definito γ come l'incremento ad ogni passo

- passo 2) $\beta(-a+2) = \beta(-a+1) + \beta(-a+2) - \beta(-a+1) + \dfrac{1-p}{p}\gamma = \gamma + \gamma\dfrac{1-p}{p}$
- passo 3)
  $\beta(-a+3) = \beta(-a+2)+\beta(-a+3)-\beta(-a+2) = \beta(-a+2)+\dfrac{1-p}{p}[\beta(-a+2)-\beta(-a+1)] =$
  $\gamma + \gamma\dfrac{1-p}{p} + \gamma(\dfrac{1-p}{p})^2$
- ...
- $\beta(x) = \gamma\Big[1 + \dfrac{1-p}{p} + \Big(\dfrac{1-p}{p}\Big)^2 + \Big(\dfrac{1-p}{p}\Big)^3 ... + \Big(\dfrac{1-p}{p}\Big)^{x+a-1}\Big]$

Quindi si ha che

$$\beta(x) = \gamma \cdot \sum_{k=0}^{x+a-1}\Big(\frac{1-p}{p}\Big)^k = \gamma \cdot \frac{1-\Big(\frac{1-p}{p}\Big)^{x+a}}{1-\frac{1-p}{p}} \tag{36}$$

Per trovare γ, poniamo :

$$\beta(b) = 1 = \gamma \cdot \frac{1-\Big(\frac{1-p}{p}\Big)^{b+a}}{1-\frac{1-p}{p}} \implies \gamma = \frac{1-\frac{1-p}{p}}{1-\Big(\frac{1-p}{p}\Big)^{b+a}} \tag{37}$$

Quindi, in forma esplicita :

$$\beta(x) = \frac{1-\Big(\frac{1-p}{p}\Big)^{x+a}}{1-\Big(\frac{1-p}{p}\Big)^{b+a}} \tag{38}$$

Osservazione : la probabilità che il gioco non finisca mai è 0, infatti presi i due eventi della rovina di A e di B, si ha che :

$$\alpha(a) + \beta(b) = \frac{1-\Big(\frac{1-p}{p}\Big)^{x+b}}{1-\Big(\frac{1-p}{p}\Big)^{b+a}} + \frac{1-\Big(\frac{1-p}{p}\Big)^{x+a}}{1-\Big(\frac{1-p}{p}\Big)^{b+a}} = 1 \tag{39}$$

La somma dei due eventi disgiunti è 1, quindi, necessariamente si causerà la rovina di uno dei due giocatori.

Osservazione : Se $p = \dfrac{1}{2}$, si ha :

$$\lim_{p \to \frac{1}{2}} \beta(x) = \lim_{p \to \frac{1}{2}} \frac{1 - \left(\dfrac{1-p}{p}\right)^{x+a}}{1 - \left(\dfrac{1-p}{p}\right)^{b+a}} = \frac{a+x}{a+b} \tag{40}$$

Esempio : Se A parte da 5 euro, ed ha probabilità di vittoria uguali a 0.4, e B parte da 7 euro, ed ha probabilità di vittoria uguali a 0.6, si ha che :

- La probabilità che B si rovini è : $\dfrac{1 - \left(\dfrac{1-0.4}{0.4}\right)^{5}}{1 - \left(\dfrac{1-0.4}{0.4}\right)^{5+7}} \simeq 0,05$

4 Variabili Aleatorie

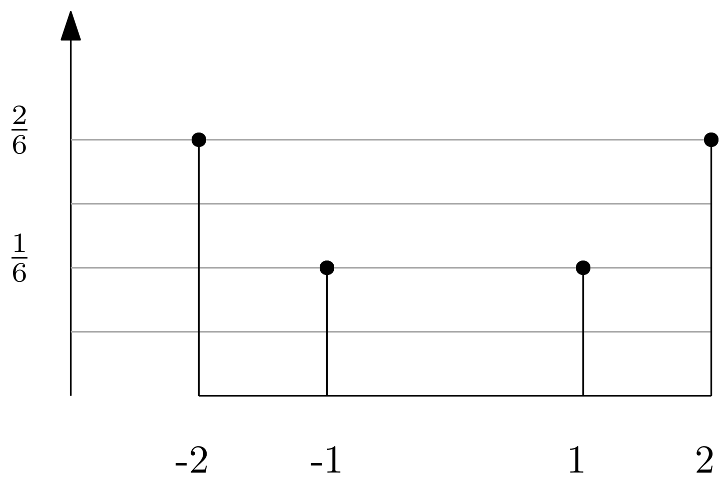
Dato uno schema probabilistico $(\Omega, \mathbb{P})$, una **variabile aleatoria** X è una funzione su Ω a valori reali:

$$X : \Omega \to \mathbb{R}$$

Tale variabile, rappresenta una "scommessa" sull'esito di un esperimento, ad ogni evento elementare o insieme di eventi elementari, è associato un valore reale.
Esempio : Si lancia un dado equo, e ci sono 2 giocatori, A e B. Se il dado da come esito 1 o 2, il giocatore A paga 2 euro, se da come esito 3, il giocatore A paga 1 euro, se da come esito 4, il giocatore B paga 1 euro, se da come esito 5 o 6, il giocatore B paga 2 euro, tale "pagamento" è esprimibile come variabile aleatoria, dove il suo valore rappresenta la quota vinta o persa da A :

$$X_1 = \begin{cases} -2 \text{ se } \omega = 1, 2 \\ -1 \text{ se } \omega = 3 \\ 1 \text{ se } \omega = 4 \\ 2 \text{ se } \omega = 5, 6 \end{cases}$$

In questo caso, X_1 non è iniettiva, se lo fosse però, in base al risultato della scommessa, sarebbe possibile conoscere l'esito del dado (in questo caso, se A paga 2 euro, non sappiamo se il dado abbia dato come esito 1 o 2).
Una variabile aleatoria X può essere rappresentata come un *istogramma*, dove sull'asse delle ascisse è presente l'immagine di X, e sulle ordinate la probabilità che ogni valore di X si verifichi. Per l'esempio preso in considerazione prima :



La probabilità $\mathbb{P}$ su Ω, *induce* una probabilità sull'immagine della variabile aleatoria, denominata $Im(X)$, mediante la funzione $\Im$:

$$\text{Sia } x_1 \in Im(X),\ \Im(\{x_1\}) := \mathbb{P}(\{\omega \in \Omega | X(\omega) = x_1\})$$

Rappresenta la probabilità che X assuma un certo valore x_1, definisce la **distribuzione** della variabile aleatoria.
Notazione :$\Im(\{x_1\})$ può essere scritto anche come $\mathbb{P}(X = x_1)$

4.1 Valore Atteso

Il **valore atteso** di una variabile aleatoria X è una funzione $\mathbb{E}$ definita nel seguente modo :

$$\mathbb{E}(X) = \sum_{x \in Im(x)} x \cdot \Im(\{x\})$$

È dato dalla somma dei possibili valori di tale variabile, ciascuno moltiplicato per la probabilità di essere assunto (ossia di verificarsi).
Indica il valore medio che X può assumere, a seconda delle probabilità di ogni valore dell'immagine di X.

Esempio : Si lancia un dado, ci sono due giocatori A e B, la variabile aleatoria X rappresenta la quota in euro vinta da A in base all'esito del dado (se negativa, è la quota che A deve pagare a B).

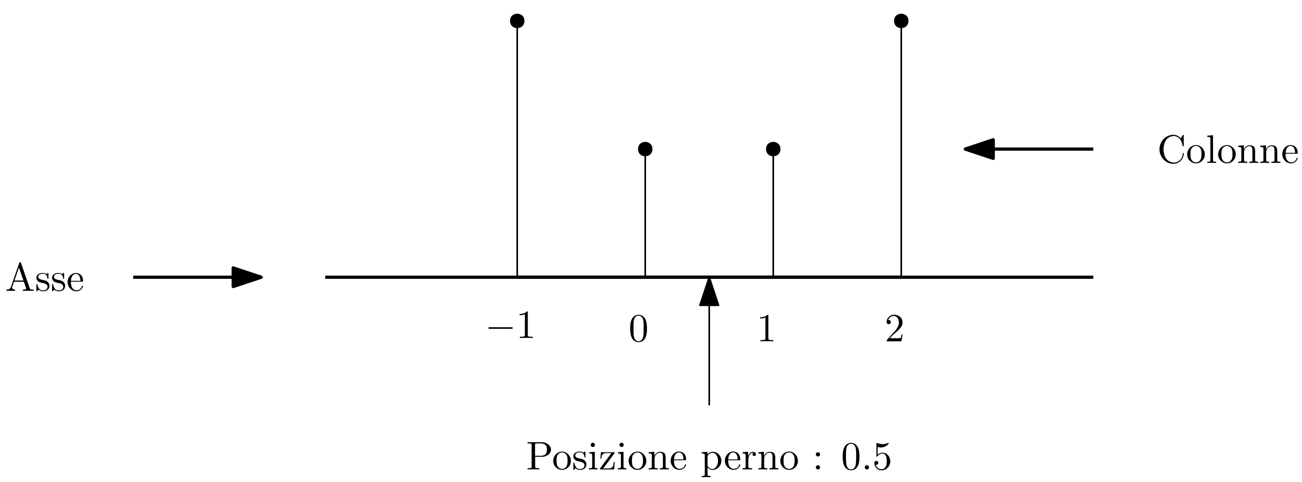
$$X = \begin{cases} -1 \text{ se } \omega = 1, 2 \\ 0 \text{ se } \omega = 3 \\ 1 \text{ se } \omega = 4 \\ 2 \text{ se } \omega = 5, 6 \end{cases}$$

In questo gioco è ovvio che A sia avvantaggiato, dato che in soli 2 esiti su 6 dovrà pagare. In base a questo schema, A sicuramente si troverà a vincere più denaro di B, ma quanto precisamente? Tale quota, è data dal valore atteso :

$$\mathbb{E}(X) = (-1) \cdot \frac{2}{6} + 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{2}{6} = 0.5$$

Quindi, deduciamo che per rendere il gioco equo, A prima di ogni lancio, dovrebbe dare a B 0.5 euro.

Una possibile rappresentazione "fisica" che si può dare al valore, atteso è la seguente :
Si consideri l'istogramma di una variabile aleatoria, si immagini l'asse delle ascisse come un asse privo di massa che si appoggia su un perno, e i valori dell'immagine della variabile, come delle colonne che hanno massa uguale alla loro probabilità, il valore atteso non è altro che, la posizione in cui si deve trovare il perno per far si che l'asse sia in equilibrio, per l'esempio precedente :



4.1.1 Linearità del Valore Atteso

Se Ω è un insieme finito, e $\mathcal{V}$ l'insieme delle sue variabili aleatorie, ossia tutte le funzioni $X : \Omega \to \mathbb{R}$, $\mathcal{V}$ assume una naturale struttura di *spazio vettoriale*.

- $\alpha \in \mathbb{R}$ ed $X \in \mathcal{V}$ si ha $(\alpha \cdot X)(\omega) = \alpha \cdot X(\omega) \qquad \forall \omega \in \Omega$
- $X, Y \in \mathcal{V}$ si ha $(X + Y)(\omega) = X(\omega) + Y(\omega) \qquad \forall \omega \in \Omega$

Effettivamente, se $|\Omega| = n$, c'è un isomorfismo canonico fra $\mathcal{V}$ ed $\mathbb{R}^n$.
Il valore atteso, definito come $\mathbb{E} : \mathcal{V} \to \mathbb{R}$ è un **applicazione lineare** :

$$\forall \alpha \in \mathbb{R}, \forall X, Y \in \mathcal{V} : \begin{cases} \mathbb{E}(\alpha \cdot X) = \alpha \cdot \mathbb{E}(X) \\ \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) \end{cases} \tag{41}$$

Dimostrazione : Si consideri la seguente osservazione preliminare :

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega | X(\omega) = x} \mathbb{P}(\{\omega\}) \cdot \sum_{x \in Im(X)} x = \sum_{x \in Im(x)} x \cdot \Im(\{x\})$$

Con la nuova definizione alternativa di valore atteso $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$ Posso dimostrare la linearità di esso :

$$\mathbb{E}(\alpha X) = \sum_{\omega \in \Omega} (\alpha X)(\omega) \cdot \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} \alpha \cdot X(\omega) \cdot \mathbb{P}(\{\omega\}) = \alpha \mathbb{E}(X) \tag{42}$$

$$\mathbb{E}(X + Y) = \sum_{\omega \in \Omega} (X + Y)(\omega) \cdot \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \cdot \mathbb{P}(\{\omega\}) = \mathbb{E}(X) + \mathbb{E}(Y) \tag{43}$$

$\blacksquare$

4.2 Varianza

Introduciamo un altro valore associato alle variabili aleatorie.
La **varianza** di una variabile aleatoria è una funzione $\mathbb{V}$ strettamente maggiore di 0, definita nel seguente modo :

$$\mathbb{V}(X) = \sum_{x \in Im(X)} (x - \mathbb{E}(X))^2 \cdot \Im(\{x\})$$

È lo scarto quadratico medio, ossia, una misura di quanto i valori si discostino quadraticamente rispetto al valore atteso. Una varianza piccola, indica che la variabile aleatoria è *distribuita* vicino il valore medio.

Si consideri l'esempio visto in precedenza : si lancia un dado, ci sono due giocatori A e B, la variabile aleatoria X rappresenta la quota in euro vinta da A in base all'esito del dado (se negativa, è la quota che A deve pagare a B).

$$X_1 = \begin{cases} -1 \text{ se } \omega = 1, 2 \\ 0 \text{ se } \omega = 3 \\ 1 \text{ se } \omega = 4 \\ 2 \text{ se } \omega = 5, 6 \end{cases}$$

In questo caso la varianza rappresenta il rischio che si corre, dato che, se la possibile perdita fosse molto alta, la varianza risulterebbe a sua volta elevata. In questo caso si ha :

$$\mathbb{V}(X_1) = (-1-0.5)^2 \cdot \frac{2}{6} + (0-0.5)^2 \cdot \frac{1}{6} + (1-0.5)^2 \cdot \frac{1}{6} + (2-0.5)^2 \cdot \frac{2}{6} \simeq 1,58 \tag{44}$$

De facto, la varianza non è elevata, in quanto il rischio di perdita, rispetto a quello di vincita, non rappresenta una grande differenza di denaro, se dovessi infatti cambiare il denaro perso si avrebbe :

$$X_2 = \begin{cases} -50 \text{ se } \omega = 1,2 \\ 0 \text{ se } \omega = 3 \\ 1 \text{ se } \omega = 4 \\ 2 \text{ se } \omega = 5,6 \end{cases}$$

$$\mathbb{V}(X_2) = (-50-0.5)^2 \cdot \frac{2}{6} + (0-0.5)^2 \cdot \frac{1}{6} + (1-0.5)^2 \cdot \frac{1}{6} + (2-0.5)^2 \cdot \frac{2}{6} \simeq 850 \tag{45}$$

4.3 Variabili Aleatorie Note

In questo paragrafo vedremo i casi più comuni di variabili aleatorie.

4.3.1 Variabile Aleatoria Certa

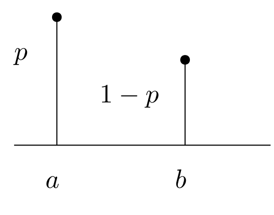
Questo è il caso in cui la variabile aleatoria X è una funzione costante:

$$\exists \tilde{x} \in \mathbb{R} | X(\omega) = \tilde{x} \forall \omega \in \Omega \text{ quindi } |Im(X)| = 1$$

Ovviamente, la varianza risulta nulla : $\mathbb{V}(X) = 0$.

4.3.2 Variabile Aleatoria di Bernoulli

Sia $p \in [0,1]$ una costante reale, e X una variabile aleatoria che ha solamente due valori nell'immagine : $Im(X) = \{a, b\}$, dove $\Im(\{a\}) = p$ e $\Im(\{b\}) = 1-p$. Dove rispettivamente $a = 1$ e $b = 0$.



In questo caso varianza e valore atteso valgono :

$$\mathbb{E}(X) = 0 \cdot (1-p) + 1 \cdot p = p \tag{46}$$

$$\mathbb{V}(X) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)(p+1-p) = p(1-p) \tag{47}$$

Se $p = \dfrac{1}{2}$, la varianza è massima.

4.3.3 Variabile Aleatoria Binomiale

Considerando lo schema di Bernoulli : $\Omega = \{0,1\}^n$, dove p è la probabilità che ω_i assuma valore 1, la variabile aleatoria binomiale X conta il numero di 1, ossia $X(\omega) = \sum_{i=1}^{n} \omega_i$.

Ad esempio, per $n = 2$ si ha :

$$X = \begin{cases} 0 \text{ se } \omega = 00 \\ 1 \text{ se } \omega = 01 \\ 1 \text{ se } \omega = 10 \\ 2 \text{ se } \omega = 11 \end{cases}$$

Abbiamo già visto in precedenza qual'è la probabilità che il numero di 1 sia k :

$$\Im(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Il valore di attesa vale :

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \cdot \Im(\{k\}) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = n \cdot p \tag{48}$$

Possiamo sfruttare la *linearità* del valore atteso 4.1.1 per calcolare il valore atteso della variabile aleatoria binomiale senza calcoli complessi. Consideriamo X_i la variabile aleatoria definita nel seguente modo :

$$X_i = \begin{cases} 1 \text{ se l'}i\text{-esimo lancio ha fatto testa} \\ 0 \text{ se l'}i\text{-esimo lancio ha fatto croce} \end{cases} \tag{49}$$

Per cui risulta $\mathbb{E}(X_i) = p$, da qui osservo che :

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} p = n \cdot p \tag{50}$$

Calcoliamo adesso la varianza della variabile aleatoria di Bernoulli, ma prima facciamo un'osservazione, in generale è vero che :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Dimostrazione :

$$\mathbb{V}(X) = \sum_{x \in Im(X)} (x - \mathbb{E}(X))^2 \cdot \Im(\{x\}) = \mathbb{E}(X - \mathbb{E}(X)^2) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2)$$

$$= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + \mathbb{E}(\mathbb{E}(X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(\mathbb{E}(X)^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \ \blacksquare$$

Quindi calcoliamo il valore di $\mathbb{E}(X^2)$ per la variabile aleatoria di bernoulli :

$$\mathbb{E}(X^2) = \sum_{k=0}^{n} k^2 \mathbb{P}(X = k) = \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \tag{51}$$

$$= \sum_{k=1}^{n} 1 \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=2}^{n-1} (k-1) \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \tag{52}$$

$$= np + \sum_{k=2}^{n} \frac{n(n-1)(n-2)!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \tag{53}$$

Sostituisco $h = k - 2$:

$$np+n(n-1)\sum_{k=2}^{n} \frac{n(n-1)(n-2)!}{(h-2)!(n-h-2)!} p^h (1-p)^{n-h-2} = np+n(n-1)\binom{n-2}{h} p^h (1-p)^{n-h-2} \tag{54}$$

$$= np + n(n-1)p^2 \tag{55}$$

Quindi :

$$\mathbb{V}(X) = \mathbb{E}(X^2) + (\mathbb{E}(X))^2 = np + n(n-1)p^2 + (np)^2 = np(1-p) \tag{56}$$

Considerando

$$X_i = \begin{cases} 1 \text{ se l'}i\text{-esimo lancio ha fatto testa} \\ 0 \text{ se l'}i\text{-esimo lancio ha fatto croce} \end{cases} \tag{57}$$

Si nota che $\mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{V}(X_i)$, sembrerebbe quindi che la varianza sia lineare, ma ciò è falso, dato che la varianza presenta uno scarto quadratico. Il fatto è che ciò è vero per la variabile aleatoria di bernoulli, perché le diverse X_i fra loro sono *indipendenti*.

4.4 Variabili Aleatorie Indipendenti

Siano $X, Y \in \mathcal{V}$ due variabili aleatorie, se $\forall x \in Im(X)$ e $\forall y \in Im(Y)$ vale che :

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

Le due variabili aleatorie si dicono **indipendenti**.

Proposizione : Se X ed Y sono indipendenti, vale la seguente identità :

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

Detto ciò, risulta chiaro del perché, per $X =\{$Variabile Aleatoria di Bernoilli$\}$, si ha che $\mathbb{V}(X) = np(1-p)$. Vediamo adesso, dimostriamo la *proposizione*.

Dimostrazione della proposizione : Sviluppo la varianza della somma :

$$\mathbb{V}(X + Y) = \mathbb{E}([X + Y - \mathbb{E}(X + Y)]^2) = \mathbb{E}([X - \mathbb{E}(X) + Y - \mathbb{E}(Y)]^2) \tag{58}$$

$$= \mathbb{E}((X - \mathbb{E}(X))^2 + (Y - \mathbb{E}(Y))^2 + 2((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))) \tag{59}$$

$$= \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \tag{60}$$

Il termine $\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$ è detto *covarianza*, bisogna dimostrare che la covarianza è nulla se X ed Y sono indipendenti.

Osservazione : Se la covarianza è nulla, allora $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$, questo perché :

$$\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \tag{61}$$

Quindi bisogna dimostrare che, se X, Y sono indipendenti, $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

Passo finale della dimostrazione :

$$\mathbb{E}(X \cdot Y) = \sum_{\omega \in \Omega} X(\omega)Y(\omega)\mathbb{P}(\omega) = \sum_{x \in Im(X)} x \cdot \sum_{y \in Im(Y)} y \cdot \sum_{\omega \in \Omega | X(\omega)=x \wedge Y(\omega)=y} \mathbb{P}(\{X = x\} \cap \{Y = y\}) \tag{62}$$

Per indipendenza : $\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$ Riscrivo :

$$\sum_{x \in Im(X)} x\mathbb{P}(X = x) \cdot \sum_{y \in Im(Y)} y\mathbb{P}(Y = y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \qquad \blacksquare \tag{63}$$

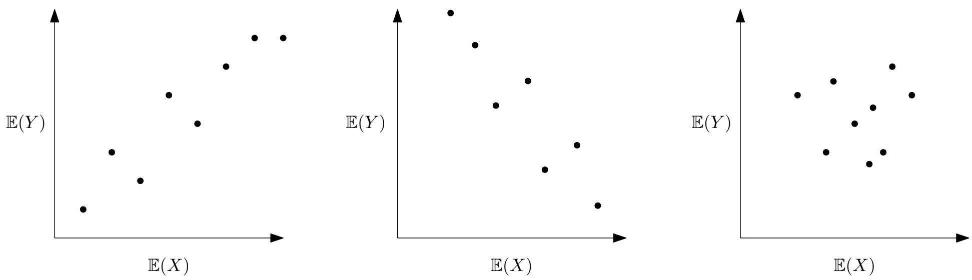
4.4.1 La Covarianza

Diamo adesso una definizione ed un significato intrenseco alla **covarianza** citata in precedenza. La covarianza è un valore associato a due variabili aleatorie ed è definita in tal modo :

$$\text{Cov}(X, Y) = (X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$$

Euristicamente parlando, la covarianza misura la relazione di dipendenza di due variabili aleatorie, può essere positiva o negativa, se $\text{Cov}(X, Y)$ è particolarmente grande, ciò significa che, sapere che X sia grande rende più probabile il fatto che Y sia grande, in generale la covarianza è limitata : $|\text{Cov}(X, Y)| \leq \sqrt{\mathbb{V}(X)} \cdot \sqrt{\mathbb{V}(Y)}$.

Esempio : Eseguo una misurazione su un circuito, X rappresenta la differenza di potenziale, ed Y l'intensità di corrente, la misurazione, avviene con una precisione tale, da poter osservare gli effetti aleatori, dovuti a fattori da noi non controllabili, per cui, nonostante la differenza di potenziale sia costante, diverse misurazioni differiscono nei valori decimali. Eseguo precisamente n misurazioni, e rappresento su un piano cartesiano il risultato, in cui $\mathbb{E}(X)$ è l'asse delle ascisse, ed $\mathbb{E}(Y)$ quello delle ordinate.



$\text{Cov}(X, Y) > 0$ — X, Y sono positivamente correlate

$\text{Cov}(X, Y) < 0$ — X, Y sono negativamente correlate

$\text{Cov}(X, Y) \simeq 0$ — X, Y non sono correlate

Se la covarianza fra due variabili aleatorie è prossima allo zero, ciò può indicare che esse sono indipendenti.

Vediamo il caso generale di variabili aleatorie indipendenti : siano $(\Omega, \mathbb{P})$ uno schema probabilistico, su cui sono dedinite le variabili aleatorie $X_1, X_2 \ldots, X_n$, diremo che esse sono indipendenti se, preso un qualsiasi punto delle loro immagini : $x_1 \in Im(X_1) \ldots, x_n \in Im(X_n)$, si ha che :

$$\mathbb{P}(X_1 = x_1, X_2 = x_2 \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \ldots, \cdot \mathbb{P}(X_n = x_n)$$

Ciò di per se, implica l'indipendenza di una qualsiasi sotto-famiglia di variabili aleatorie. Un tipico esempio da considerare è lo schema di Bernoulli, dove si ha che

$$X_i = \begin{cases} 1 \text{ se il lancio } i \text{ ha fatto testa} \\ 0 \text{ se il lancio } i \text{ ha fatto croce} \end{cases}$$

con $i = 1, 2 \ldots, n$ lanci. So che $X_1, X_2 \ldots, X_n$ sono variabili aleatorie indipendenti, quindi $\mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{V}(X_i) = n \cdot p \cdot (1-p)$.

4.5 Variabile Aleatoria Geometrica

La variabile aleatoria geometrica si comporta in modo simile alla variabile aleatoria binomiale, sia ha uno spazio di probabilità del tipo $\Omega = \{0,1\}^{\mathbb{N}}$, che significa una parola lunga infinite cifre che possono essere 0 oppure 1, la variabile aleatoria X restituisce l'i-esima cifra in cui per la prima volta compare 1, definita $X = \text{Inf}(i|\omega_i = 0)$, l'immagine di X è quindi l'insieme dei numeri naturali. La distribuzione di tale variabile aleatoria è definita nel seguente modo :

$$\mathbb{P}(X = k) = (1-p)^{k-1} \cdot p$$

Si chiama variabile aleatoria geometrica perché, con l'andare avanti dei lanci, tendendo ad infinito, la probabilità che la prima testa esca al k-esimo lancio tende a zero.

$$\lim_{k \to +\infty} (1-p)^{k-1} \cdot p = 0$$

Calcoliamo quindi la probabilità che testa, prima o poi esca :

$$\mathbb{P}(\{ \text{ prima o poi esce testa } \}) = \mathbb{P}(\bigcup_{k=1}^{+\infty} \{X = k\}) = \sum_{k=1}^{+\infty} (1-p)^{k-1} p = p \cdot \frac{1}{1-(1-p)} = 1 \quad (64)$$

Questo significa che, con l'andare avanti dei lanci, sicuramente prima o poi uscirà testa.

Il valore atteso è definito in tal modo : $\mathbb{E}(X) = \frac{1}{p}$.

Dimostrazione :

$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} k \cdot (1-p)^{k-1} \cdot p = p \cdot \sum_{k=1}^{+\infty} k \cdot (1-p)^{k-1}$$

chiamo $q = 1 - p$ ed ottengo : $p \cdot \sum_{k=1}^{+\infty} k \cdot q^{k-1}$

A questo punto, ho $k \cdot q^{k-1}$ che è la derivata di q^k, quindi riscrivo :

$$p \cdot \sum_{k=1}^{+\infty} \frac{d}{dq}(q^k) = p \cdot \frac{d}{dq}\Big(\sum_{k=1}^{+\infty} q^k\Big) = p \cdot \frac{d}{dq}\Big(\frac{1}{1-q}\Big) = p \cdot \frac{1}{(1-q)^2} = p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p} \ \blacksquare$$

4.5.1 Perdita di Memoria per la Variabile Aleatoria Geometrica

Nella variabile aleatoria geometrica, introduciamo la **funzione di sopravvivenza**, denotata e definita nel modo seguente :

$$G(n) = \mathbb{P}(\{\text{Non è uscito testa nei primi } n \text{ lanci}\}) = \mathbb{P}(X > n) = (1-p)^n$$

Dimostrazione :

$$\mathbb{P}(X > n) = \sum_{k=n+1}^{\infty} \mathbb{P}(X = k) = \sum_{k=n+1}^{\infty} p(1-p)^{k-1} \quad \text{chiamo } h = k - n - 1 \implies \tag{65}$$

$$\implies p \cdot \sum_{h=0}^{\infty} (1-p)^{n+h} = (1-p)^n \cdot \sum_{h=0}^{\infty} (1-p)^h = (1-p)^n \quad \blacksquare \tag{66}$$

Introduciamo adesso nella variabile aleatoria geometrica una condizione, ossia, so che per i primi n lanci, non è uscita testa, vogliamo sapere qual'è la probabilità che esca testa nei prossimi l lanci :

$$\mathbb{P}(X = n + l | X > n)$$

Ma in questo modello, il lancio di una moneta non influisce in alcun modo gli esiti futuri, quindi sapere che testa non è uscita, non influenza in nessun modo la probabilità che esca prossimamente :

$$\mathbb{P}(X = n + l | X > n) = \mathbb{P}(X = l)$$

Dimostrazione :

$$\mathbb{P}(X = n + l | X > n) = \frac{\mathbb{P}(X = n + l, X > n)}{\mathbb{P}(X > n)} \tag{67}$$

So che l è maggiore di zero, quindi l'evento in cui $X = n + l$, contiene anche l'evento in cui $X > n$, quindi l'intersezione di questi due ultimi è uguale al primo.

$$\frac{\mathbb{P}(X = n + l, X > n)}{\mathbb{P}(X > n)} = \frac{\mathbb{P}(X = n + l)}{\mathbb{P}(X > n)} = \frac{\mathbb{P}(X = n + l)}{G(n)} = \tag{68}$$

$$= \frac{(1-p)^{n+l-1} p}{(1-p)^n} = \frac{(1-p)^n \cdot (1-p)^{l-1} \cdot p}{(1-p)^n} = (1-p)^{l-1} \cdot p = \mathbb{P}(X = l) \quad \blacksquare \tag{69}$$

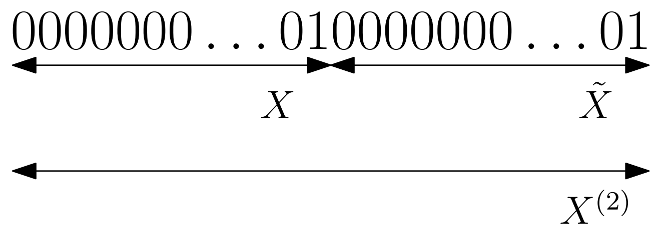
4.5.2 Variabile Aleatoria Binomiale Negativa

Adesso, si dia il caso che sia uscita testa, e che si continui a lanciare la moneta, voglio sapere la probabilità che esca testa nuovamente. Se X era la nostra variabile aleatoria geometrica, definisco una nuova variabile aleatoria :
$X^{(2)} = \{$ Lancio in cui è uscito testa la seconda volta $\} = \text{Inf}(i | \omega_i = 1 \land i > X)$, sicuramente, ciò non può che avvenire dal secondo lancio : $Im(X^{(2)}) \geq 2$, La distribuzione di tale variabile aleatoria è definita nel seguente modo :

$$\mathbb{P}(X^{(2)} = k) = (k-1)p^2(1-p)^{k-2}$$

Calcoliamoci adesso il valore atteso, sfruttando la linearità di esso, infatti, definisco una nuova variabile aleatoria $\tilde{X}$, il cui valore è uguale al numero di lanci fra la prima e la seconda testa : Risulta chiaro che, $\tilde{X}$ sia una variabile aleatoria geometrica, e che la nuova variabile $X^{(2)}$, sia la somma di $\tilde{X}$ ed X, quindi, per linearità :

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



Verifico adesso che X ed $\tilde{X}$ siano indipendenti :

$$\mathbb{P}(X = k, \tilde{X} = h) = \mathbb{P}(\{\omega_i \in \Omega | \omega_1 \dots \omega_{k-1} = 0 \wedge \omega_k = 1 \wedge \omega_{k+1} \dots \omega_{k+h-1} = 0 \wedge \omega_{k+h} = 1\}) \quad (70)$$

Essendo tutti i lanci indipendenti fra loro :

$$= (1-p)^{k-1} \cdot p \cdot (1-p)^{h-1} \cdot p = \mathbb{P}(X = k) \cdot \mathbb{P}(\tilde{X} = h) \qquad \blacksquare \quad (71)$$

Calcolando la varianza, ho che $\mathbb{V}(X^{(2)}) = \mathbb{V}(X + \tilde{X})$, avendo verificato che sono indipendenti, posso dire che $\mathbb{V}(X + \tilde{X}) = \mathbb{V}(X) + \mathbb{V}(\tilde{X})$.

Sappiamo già che $\mathbb{P}(X^{(2)} = k) = (k-1)p^2(1-p)^{k-2}$, vogliamo però calcolarlo sapendo che è la somma di 2 variabili aleatorie geometriche X e $\tilde{X}$. Vediamo però un caso generale :

Siano X ed Y due variabili aleatorie su uno spazio $(\Omega, \mathbb{P})$, definisco una nuova variabile aleatoria $Z = X + Y$. In generale, calcolare la distribuzione di Z, è difficile, a meno che non ci sia il presupposto che X ed Y siano *indipendenti*.

Calcoliamo quindi $\mathbb{P}(Z = z)$, parto prima dal calcolarmi $\mathbb{P}(Z = z, X = x)$, ho che

$$\sum_{x \in Im(X)} \mathbb{P}(Z = z, X = x) = \sum_{x \in Im(X)} \mathbb{P}(x + Y = z, X = x) = \sum_{x \in Im(X)} \mathbb{P}(Y = z - x, X = x) \forall z \in Im(Z) \quad (72)$$

Ritornando quindi al nostro esempio dei lanci di moneta :

$$\mathbb{P}(X^{(2)} = k) = \sum_{h=1}^{\infty} \mathbb{P}(X = h)\mathbb{P}(\tilde{X} = k - h) = \sum_{h=1}^{\infty} (1-p)^{h-1} p (1-p)^{k-h-1} p \quad (73)$$

Ma tale somma non può andare fino all'infinito, ma fino a $k - 1$, dato che, se
$k = h \implies \mathbb{P}(\tilde{X} = k - h) = \mathbb{P}(\tilde{X} = 0) = 0$! Allora ho :

$$\sum_{h=1}^{k-1} (1-p)^{h-1} p (1-p)^{k-h-1} p = (k-1)(1-p)^{k-2} \cdot p^2 \qquad \blacksquare \quad (74)$$

Generalizziamo adesso la variabile $X^{(2)}$, e definiamo $X^{(h)}$ come la variabile aleatoria che assume valore uguale all'indice del lancio h-esimo in cui è uscito 1, essa è denominata **variabile aleatoria negativa**.

$$X^{(h)} = \text{Inf}(\{i | i > X^{(h-1)} \wedge \omega_i = 1\})$$

La sua distribuzione vale :

$$\mathbb{P}(X^{(h)} = k) =$$
$$\mathbb{P}(\{\text{Ci sono state } h-1 \text{ teste nei primi } k-1 \text{ lanci, e al } k\text{-esimo lancio è uscita testa }\})$$
$$= \mathbb{P}(\{\omega \in \Omega | \textstyle\sum_{i=1}^{k-1} \omega_i = h - 1 \wedge \omega_k = 1\}) = \binom{k-1}{h-1} \cdot p^{h-1} \cdot (1-p)^{k-h} \cdot p$$

Il suo valore atteso è calcolabile sfruttando la linearità di esso :

$$\mathbb{E}(X^{(h)}) = \sum_{i=0}^{h} \mathbb{E}(X^{(i)}) = \sum_{i=0}^{h} \frac{1}{p} = \frac{h}{p} \tag{75}$$

4.5.3 Il Problema della Macchina da Scrivere

4.5.4 Ipotesi Modellistica

Introduciamo un concetto :

> Se un qualche evento può succedere, prima o poi succederà.

Si consideri il seguente scenario, vi è una scimmia che preme a caso dei tasti su una macchina da scrivere, la probabilità di un tasto rispetto ad un altro, non è necessariamente la stessa (magari la scimmia tende a premere più frquentemente i tasti centrali), ma la probabilità che ogni singolo tasto venga premuto è sempre maggiore di zero.

La scimmia quindi, che è immortale ed instancabile, premendo tasti all'infinito, genera una sequenza infinita di caratteri, il cui dominio di ogni carattere è l'alfabeto italiano $a, b, c \dots, z$ con aggiunta di spazi e virgole, chiamiamo tale dominio $\mathcal{A}$. La *Divina Commedia* di *Dante*, non è altro che una sequenza di k caratteri (precisamente, $k \simeq 0.5 \cdot 10^6$), se quindi lo spazio degli eventi è $\mathcal{A}^{\mathbb{N}}$, ossia le sequenze di un numero di caratteri indeterminato generate dalla scimmia, sia D l'evento, in cui la scimmia, prima o poi (ad un istante temporale indeterminato) scrive la *Divina Commedia* nella sua interezza, premendo tasti a caso.

Vedremo che $\mathbb{P}(D) = 1$, ossia, che con lo scorrere del tempo, esisterà un istante temporale, in cui la scimmia scriverà l'intera *Divina Commedia.*

4.6 Variabile Aleatoria di Poisson

4.6.1 Ipotesi Modellistica

Introduciamo una nuova variabile aleatoria, con una piccola digressione modellistica. Si dia il caso che un'azienda voglia calcolare la probabilità, in un lasso di tempo determinato (in questo caso sarà unitario, ossia di 1 secondo), di arrivo dei suoi clienti in un'ipotetica coda. Sia λ il tasso degli arrivi, ossia, in un unità di tempo, la probabilità che arrivi un cliente è λ. Si divide l'intervallo di tempo in n diversi intervalli di uguale dimensione, e la probabilità che ad ogni intervallo arrivi un cliente è di $\frac{\lambda}{n}$.

Tale problema è descrivibile con una *variabile aleatoria binomial*e 4.3.3, di parametro $\frac{\lambda}{n}$, dove n è il "numero di lanci", che in questo caso rappresenta gli istanti di tempo, in cui può arrivare un cliente.

A questo punto l'azienda, vuole essere più precisa, e rendere gli intervalli di tempo ancora più piccoli, aumentando quindi il valore di n. L'azienda, vuole calcolare la probabilità che arrivi un cliente, in un **istante infinitesimale** di tempo.

4.6.2 Definizione

Sia $\lambda \in (0, +\infty)$, e sia $X^{(n)}$ una variabile aleatoria binomiale di parametro $\frac{\lambda}{n}$, la **variabile aleatoria di Poisson**, è tale variabile binomiale, con il presupposto che n tenda all'infinito.

$$\lim_{n\to+\infty} X^{(n)}$$

La distribuzione della variabile aleatoria di Poisson vale esattamente:

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

Dimostrazione :

$$\mathbb{P}(X^{(n)} = k) = \binom{n}{k}(\frac{\lambda}{n})^n \cdot (1-\frac{\lambda}{n})^{n-k} = \frac{1}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}\lambda^k(1-\frac{\lambda}{n})^{n-k} \tag{76}$$

$$= \frac{\lambda^k}{k!}\frac{n}{n}\frac{n}{n-1}\cdots\frac{n-k+1}{n}\cdot\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \tag{77}$$

Adesso faccio il limite di n che tende ad infinito :

$$\lim_{n\to+\infty}\frac{\lambda^k}{k!}\frac{n}{n}\frac{n}{n-1}\cdots\frac{n-k+1}{n}\cdot\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} = \frac{\lambda^k}{k!}\lim_{n\to+\infty}\frac{n}{n}\frac{n}{n-1}\cdots\frac{n-k+1}{n}\cdot\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \tag{78}$$

$$\frac{\lambda^k}{k!}\lim_{n\to+\infty} 1\cdot 1\cdots 1\cdot\frac{(1-\frac{\lambda}{n})^n}{(1-0)^k} = \frac{\lambda^k}{k!}\lim_{n\to+\infty}(1-\frac{\lambda}{n})^n = \frac{\lambda^k}{k!}e^{-\lambda} \qquad \blacksquare \tag{79}$$

La variabile di Poisson quindi, ha come parametro un tasso di arrivi λ, e misura la probabilità che in un istante di tempo infinitesimale arrivi un cliente in coda.

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1 \tag{80}$$

Ovviamente, il *valore atteso*, sarà proprio $E(X^{(n)}) = \lambda$:

$$\mathbb{E}(X^{(n)}) = \sum_{k=0}^{\infty} k\mathbb{P}(X^{(n)} = k) = \sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{1}{k\cdot(k-1)!}\lambda^k = \tag{81}$$

$$e^{-\lambda}\sum_{k=1}^{\infty}\frac{1}{\cdot(k-1)!}\lambda^k \text{ sotituisco } h = k-1 \implies e^{-\lambda}\sum_{h=0}^{\infty}\lambda\cdot\frac{\lambda^h}{h!} = e^{-\lambda}\lambda\sum_{h=0}^{\infty}\frac{\lambda^h}{h!} = e^{-\lambda}\lambda e^{\lambda} = \lambda \tag{82}$$

La *varianza* vale $\mathbb{V}(X^{(n)}) = \lambda$.

4.6.3 Somma di due Variabili di Poisson

Siano X_1 ed X_2 due variabili aleatorie di Poisson, indipendenti fra loro, di parametro λ_1 e λ_2, ne voglio calcolare la somma, ossia $X = X_1 + X_2$, si avrà che :

$$\mathbb{P}(X = k) = \sum_{h=0}^{+\infty} \mathbb{P}(X_1 = h)\mathbb{P}(X_2 = k - h) = \sum_{h=0}^{+\infty} \frac{\lambda_1^h}{h!} e^{-\lambda_1} \frac{\lambda_2^{h-k}}{h - k!} e^{-\lambda_2} = \tag{83}$$

$$= e^{-(\lambda_1+\lambda_2)} \frac{1}{k!} \sum_{h=0}^{+\infty} \binom{k}{h} \lambda_1^k \lambda_2^{k-h} = e^{-(\lambda_1+\lambda_2)} \cdot \frac{(\lambda_1 + \lambda_2)^k}{k!} \tag{84}$$

Ne concludiamo che la somma di due variabili aleatorie di Poisson, è ancora una variabile aleatoria di Poisson.

4.7 Vincita Media in un Gioco Equo

In questo capitolo, dimostreremo che, se un gioco è equo, la vincita media, a prescindere dalla strategia adottata nel gioco, sarà sempre 0.

Si consideri il gioco della roulette, dove non vi è lo zero, quindi la pallina può finire esclusivamente sul rosso o sul nero, rendendo il gioco equo.

Si adotta una certa strategia, al primo turno, si gioca 1 euro, in caso di vittoria, prendo il raddoppio e finisco, in caso di sconfitta, gioco la quota giocata precedentemente ma raddoppiata, gioco per n turni, è chiaro che, se dovessi vincere, la somma vinta sarà sempre di 1 euro, se dovessi perdere invece, perderei $2^n - 1$ euro. Definisco quindi la variabile aleatoria V che rappresenta la vincita, si ricordi che ad ogni turno, la probabilità di vincere o perdere è di $1/2$.

$$V = \begin{cases} 1 \text{ con probabilità } 1 - \dfrac{1}{2^n} \\ -(2^n - 1) \text{ con probabilità } \dfrac{1}{2^n} \end{cases}$$

La probabilità di vittoria, è molto alta, e la quota vinta è bassa, la probabilità di sconfitta è bassissima, ma nel caso, si perderebbe una sommma esponenziale di denaro rispetto ai lanci fatti. Calcoliamo il valore atteso di V, che rappresenta la vincita media :

$$\mathbb{E}(V) = (1 - \frac{1}{2^n}) - (2^n - 1) \cdot \frac{1}{2^n} = 0 \tag{85}$$

Vedremo adesso che, qualunque sia la strategia adottata, la vincita media sarà sempre zero.

Introduco $X^{(i)}$, che restituisce 1 se all'i-esimo turno è uscito rosso, altrimenti 0, è ovvio che $\forall i, j \in \{1, 2 \ldots, n\}$ $X^{(i)}$ e $X^{(j)}$ sono indipendenti. Generalizziamo l'idea di strategia adottata. Al primo turno, punto sul rosso una quantità f_0, se $f_0 < 0$, vuol dire che sto puntando sul nero una quantità $|f_0|$.

Chiamo V_1 la vincita al primot turno, ed è ovvio che varrà $V_1 = f_0 \cdot X^{(1)}$

Al secondo turno, la quota da scommettere, sarà influenzata dal risultato, sarà quindi una funzione di parametro $X^{(1)}$, ossia $f_1(X^{(1)})$, la prossima vincita sarà $V_2 = V_1 + f_1(X_1) \cdot X_2$.

Ad ogni k-esimo turno, vi sarà una funzione $f_k(X_1, X_2 \ldots, X_{k-1})$, che deciderà la somma da puntare in base ai risultati dei turni precedenti, su n turni, si avrà che :

$$V_n = \sum_{i=1}^{n} f_{i-1}(X_1 \ldots, X_{i-1}) \cdot X_i$$

L'insieme di queste funzioni f_i arbitrarie, stabilisce una strategia di gioco. Possiamo calcolare il valore atteso per linearità :

$$\mathbb{E}(V_n) = \sum_{i=1}^{n} \mathbb{E}(f_{i-1}(X_1 \ldots, X_{i-1}) \cdot X_i) = \text{ per indipendenza } = \sum_{i=1}^{n} \mathbb{E}(f_{i-1}(X_1 \ldots, X_{i-1})) \cdot \mathbb{E}(X_i) \tag{86}$$

Ma per ogni $i \in \{1, 2 \ldots, n\}$, si ha che $\mathbb{E}(X_i) = 0$, ne consegue che

$$\sum_{i=1}^{n} \mathbb{E}(f_{i-1}(X_1 \ldots, X_{i-1})) \cdot \mathbb{E}(X_i) = \sum_{i=1}^{n} \mathbb{E}(f_{i-1}(X_1 \ldots, X_{i-1})) \cdot 0 = \sum_{i=1}^{n} 0 = 0 \quad \blacksquare \tag{87}$$

4.8 Legge dei Grandi Numeri

Prima di enunciare il teorema della *legge dei grandi numeri*, si consideri il seguente esempio, vi è una variabile aleatoria binomiale, che denoteremo S_n, si effettuano n lanci ed ha parametro p, vogliamo vedere cosa succede, se il numero di lanci n, tende all'infinito (da non confondere con la variabile aleatoria di Poisson, dato che il parametro in questo caso resta fissato).

Piuttosto che considerare S_n, considero $\frac{S_n}{n}$, si avrà che :

$$\lim_{n \to +\infty} \mathbb{P}(\frac{S_n}{n} = \frac{k}{n}) = 0$$

La probabilità che la variabile aleatoria assuma uno specifico valore, con l'aumentare dei lanci verso l'infinito tende a 0. Si consideri però ora il valore atteso $\mathbb{E}(\frac{S_n}{n})$, che è uguale a p (il valore atteso della variabile aleatoria binomiale è np), vedremo che, in un intorno di ampiezza 2δ del valore atteso, la probabilità che S_n assuma valore contenuto in quell'intorno, sarà certa.

$$\lim_{n \to +\infty} \mathbb{P}(p - \delta \leq \frac{S_n}{n} < p + \delta) = 1$$

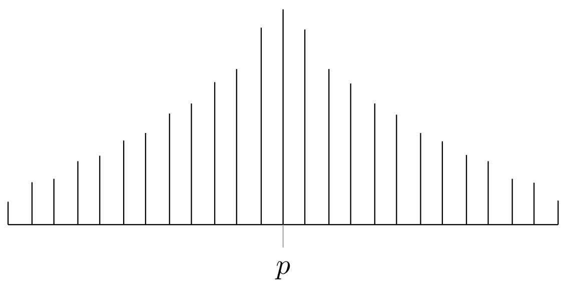
Graficamente, nell'istogramma, se il numero di lanci tende all'infinito, le stanghette al di fuori di un intorno del valore atteso tenderanno a diventare sempre più piccole, tendendo a 0. Possiamo adesso esporre il caso generale.

4.8.1 Enunciato

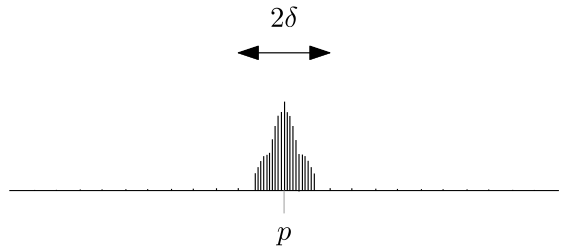
Siano $X_1, X_2 \ldots, X_n$, delle variabili aleatorie indipendenti ed identicamente distribuite (hanno la stessa distribuzione), e sia $S_n = \sum_{i=1}^{n} X_i$, si ha che :

$$\lim_{n \to +\infty} \frac{S_n}{n} = \mathbb{E}(X_i) \quad \forall i \text{ dato che la distribuzione è identica}$$

Istogramma di $\frac{S_n}{n}$



Istogramma di $\frac{S_n}{n}$
se n tende all'infinito



E ne consegue che :

$$\lim_{n\to+\infty}\mathbb{P}(|\frac{S_n}{n}-\mathbb{E}(X_i)|<\delta)=1 \quad \forall\delta>0$$

Prima di poter dimostrare tale enunciato, è necessario introdurre due concetti.

Lemma : Sia Y una variabile aleatoria arbitraria ma sempre positiva, e sia λ un numero reale sempre positivo, si ha che :

$$\mathbb{P}(Y\geq\lambda)\leq\frac{1}{\lambda}\mathbb{E}(Y)$$

Dimostrazione Lemma :

$$\mathbb{P}(Y\geq\lambda)=\sum_{y\in Im(Y)|y>\lambda}1\cdot\mathbb{P}(Y=y)\leq\sum_{y\in Im(Y)|y>\lambda}\frac{y}{\lambda}\cdot\mathbb{P}(Y=y)\leq \tag{88}$$

$$\leq\frac{1}{\lambda}\sum_{y\in Im(Y)}y\cdot\mathbb{P}(Y=y)=\frac{1}{\lambda}\mathbb{E}(Y) \quad \blacksquare \tag{89}$$

Lemma (Chebyshev) : Sia X una variabile aleatoria, e λ un numero reale sempre positivo, allora, si ha che :

$$\mathbb{P}(|X-\mathbb{E}(X)|\geq\lambda)\leq\frac{1}{\lambda^2}\cdot\mathbb{V}(X)$$

Questo vuol dire che, la probabilità che X ed il suo valore atteso differiscano al più di λ, è sempre minore o uguale a $\frac{1}{\lambda^2}\cdot\mathbb{V}(X)$.

Dimostrazione Lemma (Chebyshev) : Sia X la nostra variabile aleatoria, definisco una nuova variabile aleatoria $Y=(X-\mathbb{E}(X))^2$, noto che Y è sempre maggiore di 0, ma anche λ lo è, quindi posso applicare il **lemma** visto precedentemente e dire che :

$$\mathbb{P}(Y\geq\lambda)\leq\frac{1}{\lambda}\mathbb{E}(Y)\implies\mathbb{P}((X-\mathbb{E}(X))^2\geq\lambda)\leq\frac{1}{\lambda^2}\mathbb{E}((X-\mathbb{E}(X))^2) \tag{90}$$

Ma $(X-\mathbb{E}(X))$ è $\mathbb{V}(X)$, ne concludo che : $\mathbb{P}(|X-\mathbb{E}(X)|\geq\lambda)\leq\frac{1}{\lambda^2}\mathbb{V}(X) \quad \blacksquare$
Adesso che si hanno a disposizione questi due lemmi, è possibile procedere.

4.8.2 Dimostrazione Legge dei Grandi Numeri

Calcoliamo prima di tutto il valore atteso di $\frac{S_n}{n}$:

$$\mathbb{E}(\frac{S_n}{n}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\cdot(n\cdot\mathbb{E}(X_i)) = \mathbb{E}(X_i) \tag{91}$$

Quindi il valore atteso, è identico al valore atteso di una qualsiasi delle variabili X_i che la compone. Calcoliamo adesso la varianza :

$$\mathbb{V}(\frac{S_n}{n}) = \frac{1}{n^2}\mathbb{V}(S_n) = \text{ uso l'indipendenza } = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{1}{n^2}n\cdot\mathbb{V}(X_i) = \frac{1}{n}\mathbb{V}(X_i) \tag{92}$$

Possiamo affermare che : $\lim_{n\to+\infty}\mathbb{V}(\frac{S_n}{n}) = \lim_{n\to+\infty}\frac{1}{n}\mathbb{V}(X_i) = 0$. Voglio dimostrare che $\lim_{n\to+\infty}\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(X_i)| < \delta) = 1 \quad \forall\delta > 0$, allora, mi basta dimostrare che il suo complementare abbia probabilità 0. Passo al complementare, ossia :

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(X_i)| \geq \delta)$$

So che $\mathbb{E}(\frac{S_n}{n}) = \mathbb{E}(X_i)$, quindi riscrivo :

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| \geq \delta)$$

Ma questo, è proprio il caso del *Lemma di Chebyshev*, quindi so che :

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| \geq \delta) \leq \frac{1}{\lambda^2}\mathbb{V}(\frac{S_n}{n})$$

Adesso, voglio sapere cosa succede se n tende all'infinito, ho dimostrato prima che, se $\lim_{n\to+\infty}\mathbb{V}(\frac{S_n}{n}) = 0$, quindi posso riscrivere :

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| \geq \delta) \leq 0$$

Ma la probabilità non può essere minore di zero, quindi è certo che :

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| \geq \delta) = 0$$

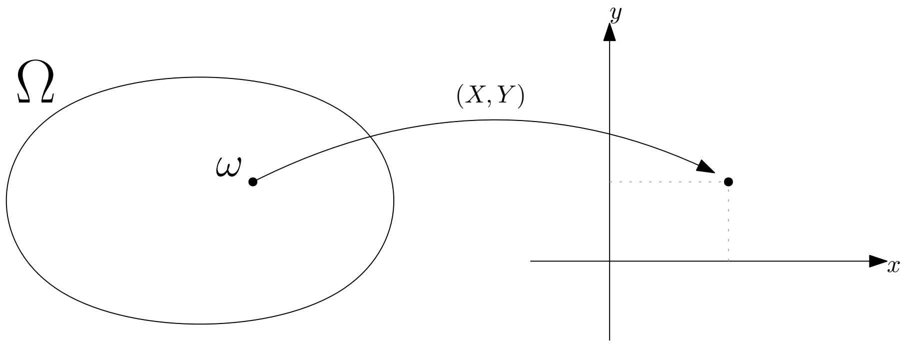
Essendo questo il complementare di $\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| < \delta)$, ne concludo che :

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| < \delta) = 1 - \mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| \geq \delta) = 1 - 0 = 1$$

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})| < \delta) = 0 \qquad \blacksquare$$

4.9 Distribuzione Congiunta di Variabili Aleatorie

Siano X ed Y due variabili aleatorie, si vuole guardare alla distribuzione di esse in maniera congiunta, ad ogni evento, sarà quindi associato il risultato delle due variabili, che sarà ovviamente un punto su $\mathbb{R}^2$:



Esempio : Si vuole lanciare un dado equo, sappiamo che $\Omega = \{1, 2, 3, 4, 5, 6\}$ e che la probabilità è uniforme. Considero due variabili aleatorie, la cui immagine rappresenta la somma in denaro vinta (o persa, se negativa) in base all'esito del dado :

$$X = \begin{cases} -1 \text{ se } \omega = 1, 2 \\ 0 \text{ se } \omega = 3, 4 \\ 1 \text{ se } \omega = 5, 6 \end{cases} \qquad Y = \begin{cases} -2 \text{ se } \omega = 1, 2, 3 \\ 2 \text{ se } \omega = 4, 5, 6 \end{cases}$$

Posso considerare la probabilità congiunta, ossia la probabilità che X assuma un certo valore, ed Y ne assuma un altro, rappresento una tabella, in cui sulle righe considero l'immagine di X e sulle colonne l'immagine di Y, tale tabella sarà riempita con le rispettiva probabilità :

$Y \backslash X$	-1	0	1
-2	$2/6$	$1/6$	0
2	0	$1/6$	$2/6$

Si noti come, ovviamente, possono esserci degli esiti congiunti che hanno probabilità 0, se il dado da come esito 1, la variabile X non può essere uguale ad 1 se la variabile Y è uguale a -2, in quanto quest'ultima, restringe i possibili esiti del dado a 1,2 o 3, e per questi esiti, la variabile X non può valere 1.

Definizione : La distribuzione congiunta delle variabili aleatorie X ed Y è data da :

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(\{\omega \in \Omega | X(\omega) = x \wedge Y(\omega) = y\})$$

Osservazione : Da una distribuzione congiunta è possibile ricavare le distribuzioni delle variabili che la compongono, ed esse sono dette *distribuzioni marginali*, dalla tabella, sommando tutti i valori di una riga, è possibile ottenere la distribuzione di Y, e sommando una colonna di X :

kk	-1	0	1	
-2	$2/6$	$1/6$	0	$2/6 + 1/6 = \boxed{3/6}$
2	0	$1/6$	$2/6$	$2/6 + 1/6 = \boxed{3/6}$
	$2/6+$ $0 =$	$1/6+$ $1/6 =$	$2/6+$ $0 =$	
	$2/6$	$2/6$	$2/6$	

Si ha logicamente che :

$$\mathbb{P}(X=x_0)=\sum_{y\in Im(Y)}\mathbb{P}(X=x_0,Y=y)\qquad \mathbb{P}(Y=y_0)=\sum_{x\in Im(X)}\mathbb{P}(X=x,Y=y_0)$$

Se la probabilità congiunta è uguale al prodotto delle distribuzioni marginali, allora X ed Y sono indipendenti.

L'esito di una variabile aleatoria, può ovviamente influenzare la probabilità che un altra variabile aleatoria dia un certo risultato, considerando l'esempio precedente, sapere che la variabile Y ha assundo valore uguale a 2, rende nulla la probabilità che X possa assumere valore -1. In generale :

$$\mathbb{P}(Y=y|X=x)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(X=x)}$$

4.10 Variabile Aleatoria Multinomiale

Vediamo subito un esempio di variabili aleatorie congiunte, introducendo una nuova variabile aleatoria, ossia quella *multinomiale* : Vi è un eseprimento non binario, che ha k possibili esiti, ossia $\{1,2,3\ldots,k\}$, ogni esito i, ha probabilità p_i, dove ovviamente $\sum_{i=1}^{k}p_i=1$. Tale esperimento, viene ripetuto n volte. Stabilisco adesso delle nuove variabili aleatorie, $\forall i$, ho che $X_i=$ numero di volte che l'esperimento da esito i. La congiunzione di queste variabili aleatorie, sarà appunto la nostra variabile multinomiale :

$$(X_1,X_2\ldots,X_k)\sim\text{Mult}(n,k,\{p_1,p_2\ldots,p_k\})$$

Ricordando che k è il numero di possibili esiti, n il numero di volte in cui si prova l'esperimento, e $\{p_1,p_2\ldots,p_k\}$ le rispettive probabilità per ogni esito. Si avrà distribuzione :

$$\mathbb{P}(X_1=n_1,X_2=n_2\ldots,X_k=n_k)=\frac{n!}{n_1!n_2!\cdot\cdots\cdot n_k!}p_1^{n_1}p_2^{n_2}\ldots\cdot p_k^{n_k}$$

Con ovviamente $\sum_{i=1}^{k}n_i=n$.

4.11 Variabili Aleatorie Continue

Fino a questo punto del corso, ci siamo occupati esclusivamente di variabili aleatorie, che avessero immagine in un sottoinsieme *numerabile* dei numeri reali. Vi sono situazioni, in cui una variabile aleatoria ha immagine **continua**, con cardinalità non numerabile. In questi casi, non avrà più senso chiedersi quale sia la probabilità che la variabile aleatoria assuma un certo valore, perché sarà sempre nulla.

$$\mathbb{P}(X=c)=0,\ \forall c\in Im(X)$$

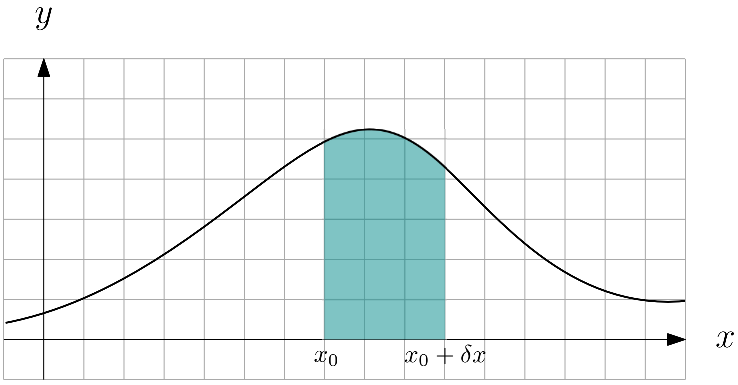
La domanda da porci sarà quindi, la probabilità che la variabile aleatoria, risulti in un certo intervallo :

$$\mathbb{P}(X\in[x,x+\delta x])=f_X(x)\cdot\delta x+o(\delta x)$$

La funzione f_X, è dettà **densità di probabilità**, e rappresenta esattamente il grafico sul piano cartesiano, dei valori che può assumere X, con le rispettive probabilità (la versione continua dell'istogramma). Quindi tale funzione descrive la distribuzione, si ha ovviamente che $f_X > 0$.

La probabilità che una variabile aleatoria continua rientri in un certo intervallo, sarà uguale alla *somma nel continuo* delle probabilità dei valori all'interno dell'intervallo. Dato che si parla di somma nel continuo, sarà essa uguale all'integrale definito nell'intervallo, della funzione densità. Geometricamente parlando, la probabilità sarà quindi uguale all'area sottesa dalla curva f_X :

Grafico di f_X



$$\mathbb{P}(X \in [x_0, x_0 + \delta x]) = \int_{x_0}^{x_0+\delta x} f_X(x)dx =$$

Ovviamente, la somma di tutte le probabilità di ogni evento, ossia l'evento certo, deve essere uguale ad 1, si avrà quindi che :

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

Il valore atteso sarà :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx$$

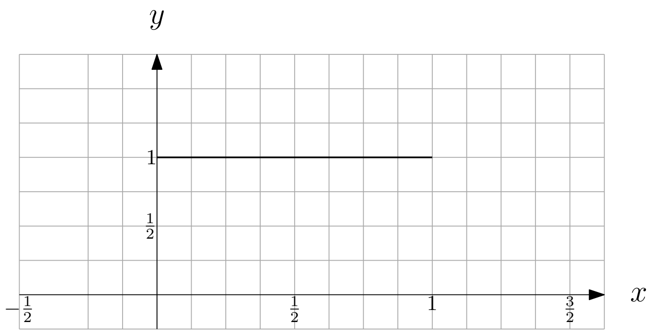
La varianza sarà sempre $\mathbb{V}(X) = \mathbb{E}(X^2) + [E(X)]^2$, calcoliamoci $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot f_X(x)dx \tag{93}$$

Esempio : Sia X una variabile aleatoria continua uniforme nell'intervallo $[0, 1]$, si può immaginare praticamente come un generatore di numeri casuali, la funzione di densità sarà definita in tal modo :

$$f_X(x) = \begin{cases} 0 \text{ se } x < 0 \\ 1 \text{ se } 0 \leq x \leq 1 \\ 0 \text{ se } x > 1 \end{cases}$$

Ed avrà grafico:



Siano $a < b$, si avrà che :

$$\mathbb{P}(a \leq x \leq b) = \int_a^b f_X(x)dx = \int_a^b 1dx = \left|x\right|_a^b = b - a$$

Con seguente valore atteso :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx = \int_{-\infty}^{0} x \cdot 0dx + \int_0^1 x \cdot 1dx + \int_0^{+\infty} x \cdot 0dx = \left|\frac{x^2}{2}\right|_0^1 = \frac{1}{2} \tag{94}$$

E varianza :

$$\mathbb{V}(X) = \int_{-\infty}^{+\infty} x^2dx + \Big[\int_{-\infty}^{+\infty} xdx\Big]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \tag{95}$$

Siano X ed Y due variabili aleatorie continue *indipendenti*, con rispettive funzioni densità f_X ed f_Y, allora, la variabile $Z = X + Y$, avrà densità $f_Z(z) = \displaystyle\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx$.

4.11.1 Distribuzione di Variabili Aleatorie Continue

Introduciamo adesso, la funzione *distribuzione* $F_X : \mathbb{R} \to [0,1]$, per una variabile aleatoria continua X, definita in tal modo :

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

Tale funzione è *crescente*, si ha che :

$$F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0 \qquad F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1$$

La funzione distribuzione, è una **primitiva** della densità :

$$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x]) = \int_{-\infty}^{x} f_X(y)dy$$

$$\frac{d}{dx}F_X(x) = f_X(x)$$

Siano X, Y due variabili aleatorie continue ed indipendenti, con rispettive funzioni densità f_X ed f_Y, definiamo $Z = \max(X, Y)$, calcoliamo la distribuzione :

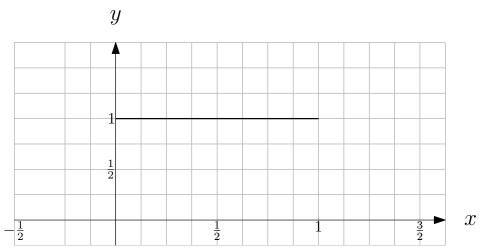
$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(\max(X,Y) < z) = \mathbb{P}(X \leq z, Y \leq z) \tag{96}$$

Per indipendenza ho :

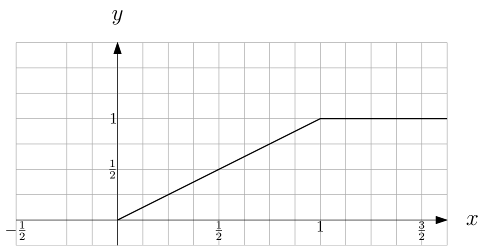
$$\mathbb{P}(X \leq z)\mathbb{P}(Y \leq z) = F_X(z)F_Y(z) \tag{97}$$

Vediamo adesso la distribuzione della variabile aleatoria continua uniforme nell'intervallo $[0, 1]$:

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(y)dy = \begin{cases} 0 \text{ se } x < 0 \\ \int_0^x 1dy \text{ se } 0 \leq x \leq 1 \\ 1 \text{ se } x > 1 \end{cases} = \begin{cases} 0 \text{ se } x < 0 \\ x \text{ se } 0 \leq x \leq 1 \\ 1 \text{ se } x > 1 \end{cases} \tag{98}$$



Ha primitiva $\longrightarrow$



Variabile Aleatoria Composta : Sia X una variabile aleatoria continua ed indipendente con densità f_X, e sia $\phi : \mathbb{R} \to \mathbb{R}$ una funzione crescente. Introduco una nuova variabile aleatoria $Y = \phi(X)$, e mi chiedo quale sia la sua densità. Inizio calcolando la probabilità che Y rientri in un qualsiasi intervallo :

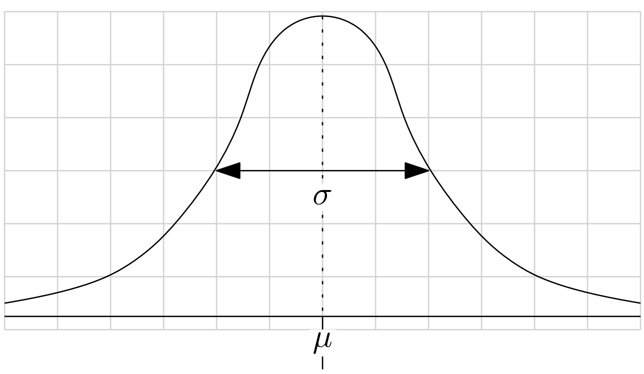
$$\mathbb{P}(Y \in [y, y + \delta y]) = \mathbb{P}(\phi(X) \in [y, y + \delta y]) \tag{99}$$

Essendo ϕ monotona crescente, riscrivo :

$$\mathbb{P}(x \in [\phi^{-1}(Y), \phi^{-1}(Y) + \delta y]) = f_X(\phi^{-1}(Y))\frac{1}{\phi(\phi^{-1}(Y))}\delta y = f_Y \tag{100}$$

4.12 Variabile Aleatoria Gaussiana

Per i processi di misurazione, è nota la teoria degli errori; Quando si misura l'intensità di corrente passante per un punto in un circuito, si legge un valore reale, soggetto a fluttuazioni totalmente aleatorie, dovute dall'inaccuratezza dello strumnento di misura, causante errori casuali, non sistematici.
Supponiamo di misurare un milione di volte tale corrente, e di disegnarne l'istogramma, esso, avrà la seguente forma :



μ è il valore "teorico" che dovrebbe assumere la misurazione (nell'esempio dell'intensità di corrente, sarebbe quello dato dalle leggi di *Ohm*), σ invece, è l'ampiezza della fluttuazione.

Questo comportamente è descritto dalla variabile aleatoria *Gaussiana* o *Normale*, definita con il simbolo $X \sim \mathcal{N}(\mu, \sigma^2)$, che accetta appunto i due parametri appena presentati (l'ampiezza della fluttuazione è sempre positiva, quindi si scrive al quadrato). Il grafico visto precedentemente non è altro che la densità di probabilitò, definito in tal modo :

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Essendo una densità, si avrà che l'area sottesa al grafico è proprio 1 :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \begin{bmatrix} z = \dfrac{x-\mu}{\sigma} \\ dz = \dfrac{dx}{\sigma} \end{bmatrix} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz = 1 \tag{101}$$

Gauss, calcolò che tale integrale è proprio uguale ad 1 (la dimostrazione non verrà presentata in questo corso).

Segue il calcolo del valore atteso :

$$\mathbb{E}(X) = x \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \begin{bmatrix} z = \dfrac{x-\mu}{\sigma} \\ dz = \dfrac{dx}{\sigma} \end{bmatrix} = \int_{-\infty}^{+\infty} (\mu+\sigma z) \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz \tag{102}$$

$$= \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz + \sigma \int_{-\infty}^{+\infty} z \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz \tag{103}$$

Il primo termine, è proprio la densità di probabilità, che vale 1, il secondo termine, è l'integrale di una funzione dispari, sappiamo quindi valere 0 :

$$\mathbb{E}(X) = \mu \cdot 1 + \sigma \cdot 0 = \mu \tag{104}$$

Segue il calcolo della varianza :

$$\mathbb{V}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \begin{bmatrix} z = \dfrac{x-\mu}{\sigma} \\ dz = \dfrac{dx}{\sigma} \end{bmatrix} = \int_{-\infty}^{+\infty} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz \tag{105}$$

$$= \sigma^2 \int_{-\infty}^{+\infty} z * z \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz \tag{106}$$

OSSERVAZIONE FONDAMENTALE : $\frac{d}{dz}(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}) = -z \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$, posso quindi integrare per parti :

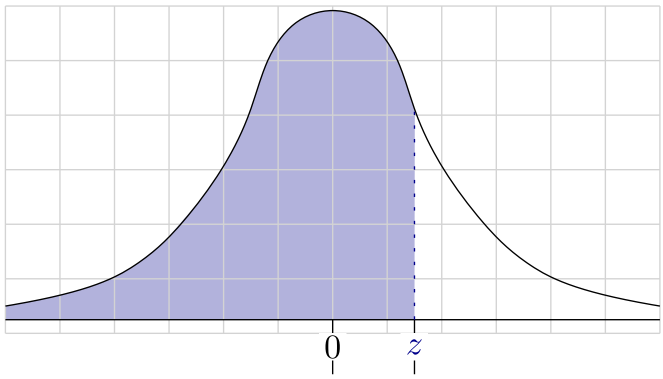
$$= \sigma^2 \left[-z \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{d}{dz} z \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz \right] = \sigma^2(0+1) = \sigma^2 \tag{107}$$

4.12.1 Standardizzazione

Una qualsiasi variabile aleatoria Gaussiana può essere "standardizzata" in modo che i suoi parametri risultino 0 per il valore teorico, ed 1 per l'ampiezza, sia $X \sim \mathcal{N}(\mu, \sigma^2)$, considero una nuova variabile aleatoria $Z := \dfrac{X-\mu}{\sigma}$, ossia, misuro X a partire dalla media in unità σ, si avrà che $Z \sim \mathcal{N}(0, 1)$.

Si può dimostrare che, l'integrale su tutta la retta reale della densità, è uguale ad 1, ma non esiste un modo esplicito per calcolare analiticamente l'integrale definito della funzione di densità, in quanto non è possibile trovare una sua primitiva, sia c una costante, ed $X \sim \mathcal{N}(\mu, \sigma^2)$, non è possibile calcolare $\mathbb{P}(X < c)$. Tramite il calcolo numerico, si può approssimare in maniera molto precisa però l'integrale definito, di fatto, esiste una tabella, che stila i valori di tale integrale, con una precisione di 4 cifre decimali. Sia ϕ una funzione integrale definita nel seguente modo :

$$\phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}\, dx = \mathbb{P}(X < z) =$$



La tabella ha sulle ascisse e sulle ordinate dei valori reali, ed alla posizione (i, j), è riportato il valore della funzione ϕ della somma dei numeri reali alle posizioni i e j :

z	0.00	0.01	0.02	...	0.04	...	0.09
0.0	0.5	0.5040	0.5080	*	*	*	*
0.1	0.5398	0.5438	0.5478	*	*	*	*
0.2	*	*	*	*	*	*	*
.	*	*	*	*	*	*	*
1.3	*	*	*	*	0.9099	*	*
.	*	*	*	*	*	*	*
3.4	*	*	*	*	*	*	0.9998

Nell'esempio riportato, $\phi(1.34) = \displaystyle\int_{-\infty}^{1.34} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}\, dx = 0.9998$.

Esempio : Sia $X \sim \mathcal{N}(-1, 4)$, si calcoli $\mathbb{P}(X > 0)$. Definisco $Z = \dfrac{X-(-1)}{\sqrt{4}} = \dfrac{X+1}{2} \implies X = 2Z - 1$. Ho che $\mathbb{P}(X > 0) = \mathbb{P}(2Z + 1 > 0) = \mathbb{P}(Z > -\frac{1}{2})$ che per simmetria della densità di probabilità è uguale a $\mathbb{P}(Z < \frac{1}{2}) = 0.6915$ consultando la tavola.

4.12.2 Teorema del Limite Centrale

Siano n variabili aleatorie indipendenti : $X_1, X_2 \dots, X_n$ identicamente distribuite, con $\mathbb{E}(X_1) = \mu$ e $\mathbb{V}(X_i) = \sigma^2$, considero la loro somma, ossia $S_n = \sum_{i=1}^{2} X_i$, abbiamo già visto nel teorema della legge dei grandi numeri che $\mathbb{E}(S_n) = n\mu$ e $\mathbb{V}(S_n) = n\sigma^2$. Considero ora una nuova variabile : $\frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}}$. Se n tende ad infinito, tale variabile aleatoria tenderà a diventare la variabile aleatoria Gaussiana normalizzata :

$$\lim_{n\to+\infty} \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}} = Z \sim \mathcal{N}(0,1)$$

$$\lim_{n\to+\infty} \mathbb{P}(a \leq \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}} \leq b) = \mathbb{P}(a \leq Z \leq b) = \phi(-a) \cdot \phi(b)$$

Tale teorema, estende ciò che diceva la legge dei grandi numeri, quando n tende ad infinito, la somma di tali variabili aleatorie, in un intorno del valore atteso, convergerà perfettamente alla funzione di densità della variabile aleatoria Gaussiana.

4.12.3 Verifica dell'Equità di una Moneta

Vediamo un caso di applicazione del teorema del limite centrale. Vi è una moneta *presunta equa*, bisogna effettuare un test statistico, per decretare se essa, è effettivamente equa oppure no. Si effettuano 10000 lanci, e si osserva l'esito dell'esperimento. L'esito, vede la moneta aver fatto testa per 5500 volte, e croce per 4500 volte.

In base a questo dato, vogliamo dare un decreto sull'equità, ci chiediamo se questa **fluttuazione** sia compatibile con il fatto che la moneta sia equa. S_n è la variabile aleatoria binomiale che restituisce il numero di teste nel lancio di una moneta equa, vogliamo calcolare $\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq 500) = \mathbb{P}(|S_n - 5000| \geq 500)$. Se la probabilità di tale evento è troppo piccola, esso sarà un chiaro segnale che la moneta non sia equa.

Utilizziamo la disuguaglianza di Chebysnev, che ricordiamo affermare che

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{1}{\lambda^2} \cdot \mathbb{V}(X)$$

Ho quindi che :

$$\mathbb{P}(|S_n - 5000| \geq 500) \leq \frac{1}{500^2} \cdot \mathbb{V}(S_n) \frac{1}{500^2} \cdot 10^4 \frac{1}{4} = \frac{1}{100}$$

So che la probabilità che escano 5500 teste, è minore di un centesimo, ma questo dato, non è abbastanza preciso, ci serve un approssimazione più specifica. Possiamo calcolare analiticamente questa probabilità, facendo tendere il numero di lanci verso infinito, come prima cosa noto che :

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq 500) = \mathbb{P}(\frac{|S_n - \mathbb{E}(S_n)|}{\sqrt{\mathbb{V}(S_n)}} \geq \frac{500}{\sqrt{\mathbb{V}(S_n)}})$$

Considero quindi la nuova variabile aleatoria $Z = \dfrac{|S_n - \mathbb{E}(S_n)|}{\sqrt{\mathbb{V}(S_n)}}$, e so che se, per il teorema del limite centrale, se $n \to +\infty$, allora Z tenderà ad essere una variabile aleatoria Gaussiana $\mathcal{N}(0,1)$. A questo punto, avrò che :

$$\lim_{n \to +\infty} \mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq 500) = \mathbb{P}(|Z| \geq 10)$$

Conosciamo i valori della probabilità che una variabile Gaussiana rientri in un certo intervallo, e consultando la tavola, vedremo che $\mathbb{P}(|Z| \geq 10) \simeq e^{-100}$. È un risultato totalmente irrisorio, e possiamo affermare con certezza che la moneta *non* sia equa.

4.13 Simulazione di Variabili Aleatorie

Sia in matematica pura che nelle applicazioni, risulta estremamente utili poter simulare una variabile aleatoria.

Consideriamo X la variabile aleatoria di Bernoulli di parametro p, voglio simulare tale variabile, utilizzando il generatore di numeri casuali, ossia la variabile aleatoria uniforme U nell'intervallo reale $[0,1]$. In questa variabile, abbiamo tutta l'*aleatorietà* necessaria per simulare X, infatti, X può essere definita come :

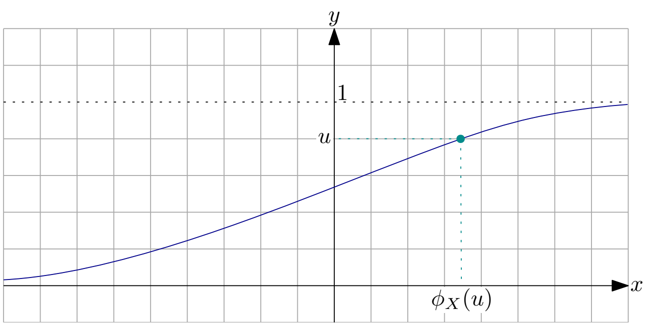
$$X = \begin{cases} 0 & \text{se} \quad 0 \leq U \leq (1-p) \\ 1 & \text{se} \quad (1-p) < U \leq p \end{cases} \quad \sim \text{V.A. di Bernoulli di parametro } p$$

Lemma :Sia X una qualsiasi variabile aleatoria arbitraria, allora esiste un applicazione $\phi_X : [0,1] \to \mathbb{R}$ tale che $X \sim \phi_X(U)$, dove U è la variabile aleatoria uniforme nell'intervallo reale $[0,1]$. Il lemma, afferma che una qualsiasi variabile aleatoria può essere simulata con una variabile aleatoria uniforme. Bisogna adesso capire, come trovare questa funzione ϕ_X.

Osservazione : La funzione ϕ_X è costruita esplicitamente a partire dalla variabile X da simulare. Essa sarà la funzione inversa della distribuzione di X.

Sia X una variabile aleatoria arbitraria continua, e sia f_X la sua densità, sempre positiva. La funzione distribuzione F_X, sappiamo essere $\int_{-\infty}^{x} f_X(t)dt$, essendo la sua derivata sempre positiva, possiamo affermare che F_X sia una funzione monotona crescente, quindi *biettiva*.

$$F_X : (-\infty, +\infty) \to [0,1] \quad \text{pongo} \quad \phi_X = F_X^{-1} \implies \phi_X(U) = X$$



Dimostrazione : Basta verificare che la funzione di distribuzione di $\phi_X(U)$ sia uguale a quella di X, ossia F_X.

$$\forall x \in \mathbb{R} \quad \mathbb{P}(\phi(X) \leq x) = \mathbb{P}(F_X^{-1}(U) \leq x) = \text{ essendo } F_X \text{ monotona crescente } = \mathbb{P}(U \leq F_X(x))$$

$$\mathbb{P}(U \leq F_X(x)) = F_X(x) \quad \blacksquare$$

5 Argomenti Extra

5.1 Catene di Markov

Supponiamo di voler rappresentare un generico *sistema*, che può trovarsi in differenti stati nel tempo. Il sistema, si trova ad un certo stato i al tempo n. Possiamo rappresentare lo stato del sistema come una variabile aleatoria, $X_n = i$ significa che il sistema al tempo n è nello stato i, il tempo è discreto, quindi possiamo descrivere l'andamento del sistema nel tempo con una successione di variabili aleatorie.

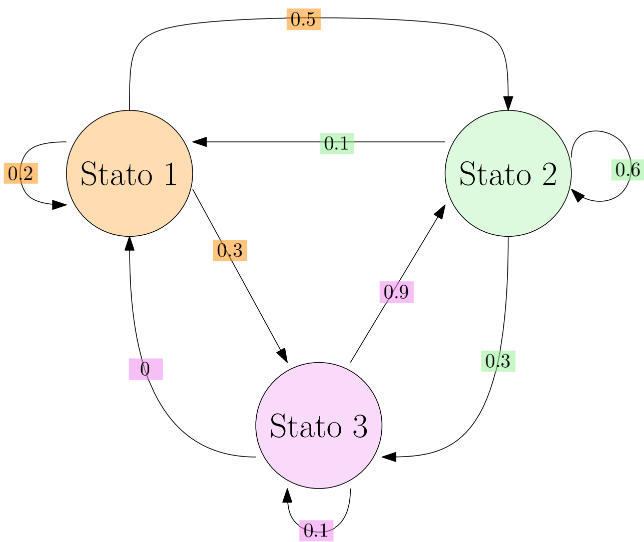
$$X_0, X_1, X_2 \ldots, X_n$$

I possibili stati che può assumere il sistema sono ben definiti in un insieme $\{0, 1, 2 \ldots, M\}$. Con lo scorrere del tempo, il sistema cambia da uno stato ad un'altro di quelli disponibili con una certa probabilità.

Diremo che P_{i_j}, è la probabilità che ha il sistema, di passare allo stato j considerando che si trova nello stato i, e considerando lo storico degli stati precedenti.

$$P_{i_j} = \mathbb{P}(X_{n+1} = j | X_n = i \wedge X_{n-1} = i_{n-1} \wedge X_{n-2} = i_{n-2} \cdots \wedge X_0 = i_0)$$

Risulta naturale rappresentare tale sistema, che chiameremo *catena di Markov*, con un grafo, in cui i nodi sono i possibili stati, ed ogni nodo i ha un arco che lo collega con tutti gli altri nodi, etichettato da un numero rapprestante la probabilità che si passi dallo stato i allo stato indicato dall'arco.



Da come si può notare, risulta ovvio che la somma di tutte le probabilità che ha uno stato di passare agli altri possibili stati deve essere 1:

$$\forall i \in \{0, 1 \ldots, M\} \quad \sum_{j=0}^{M} P_{i_j} = 1$$

Risulta altrettanto più naturale, disporre le probabilità che ogni stato ha di passare ad un altro nell'istante temporale successivo, in una matrice quadrata $M \times M$:

$$P = \begin{bmatrix} P_{0_0} & P_{0_1} & P_{0_2} & \dots & P_{0_M} \\ P_{1_0} & P_{1_1} & P_{1_2} & \dots & P_{1_M} \\ P_{2_0} & P_{2_1} & P_{2_2} & \dots & P_{2_M} \\ \vdots & \vdots & \vdots & & \vdots \\ P_{M_0} & P_{M_1} & P_{M_2} & \dots & P_{M_M} \end{bmatrix}$$

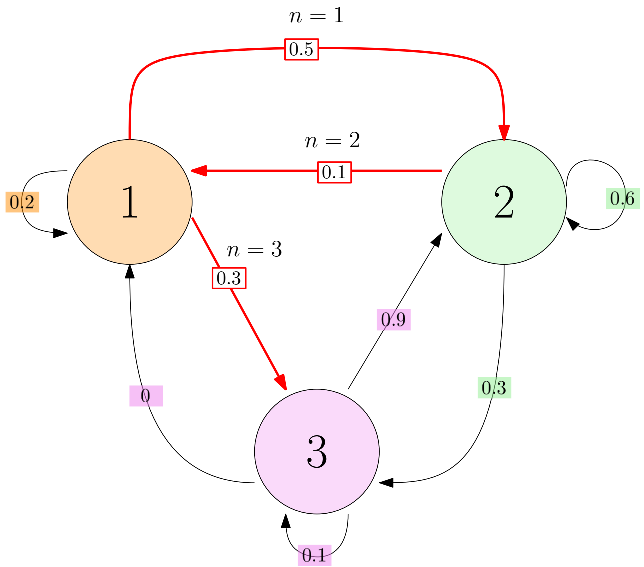
Conoscere i valori di tale matrice, è necessario per poter calcolare tutte le probabilità relative alla catena di Markov , ad esempio la probabilità che il sistema abbia seguito un certo sviluppo nel tempo. De facto, si ha che :

$$\mathbb{P}(X_n = i_n \wedge X_{n-1} = i_{n-1} \dots, X_1 = i_1 \wedge X_0 = i_0) = P_{i_{n-1} i_n} \cdot P_{i_{n-2} i_{n-1}} \dots P_{i_1 2} \cdot P_{i_0 1} \cdot \mathbb{P}(X_0 = i_0)$$

Si è scritto esplicitamente il termine $\mathbb{P}(X_0 = i_0)$, dato che all'istante di tempo zero, non si è chiaro da quale stato/nodo partire, definiremo quindi una funzione $\mu : \{0, 1 \dots, M\} \to [0, 1]$, che dato uno stato/nodo, ne restituisce la probabilità che al tempo iniziale 0, ci si trovi in quello stato/nodo. Per convenzione, se ci sono M stati/nodi, si considera la probabilità uniforme, quindi $\forall i \in \{0, 1 \dots, M\} \quad \mu(i) = \dfrac{1}{M}$.

Esempio : Si consideri il sistema rappresentato dal grafo mostrato ad inizio capitolo 5.1, si considerano 3 intervalli di tempo. Qual'è la probabilità p che, partendo dallo stato 1 :

- Al primo intervallo di tempo si passa dallo stato 1 allo stato 2
- Al secondo intervallo di tempo si passa dallo stato 2 allo stato 1
- Al terzo intervallo di tempo si passa dallo stato 1 allo stato 3



$$P_{1_2} = 0.5$$
$$P_{2_1} = 0.1$$
$$P_{1_3} = 0.3$$

$$\mathbb{P}(X_1 = 2 \wedge X_2 = 1 \wedge X_3 = 3) = 0.5 \cdot 0.1 \cdot 0.3 = 0.015$$

5.1.1 Prossime Transazioni

Consideriamo adesso un altro scenario, in cui siamo nello stato i, e vogliamo calcolare le probabilità che si passi allo stato j fra due istanti temporali successivi, si denota nel seguente modo :

$$P_{i_j}^{(2)} = \mathbb{P}(X_{m+2} = j | X_m = i)$$

Possiamo definire questaq probabilità in maniera euristica nel seguente modo, si inizia dallo stato i, e si calcolano tutti i possibili "cammini" in 2 istanti temporali, e se ne sommano le probabilità di quelli che giungono nello stato j.

$$\sum_{k=0}^{M} \mathbb{P}(X_2 = j \wedge X_1 = k \wedge X_0 = i)$$

Sviluppando tale sommatoria si noterà che :

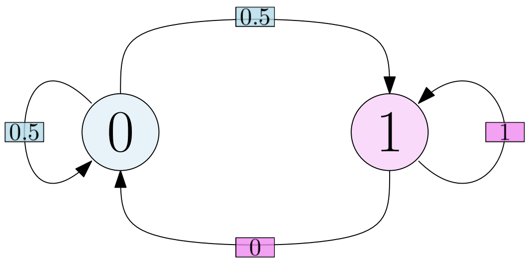
$$= \sum_{k=0}^{M} \mathbb{P}(X_2 = j | X_1 = k \wedge X_0 = i) \cdot \mathbb{P}(X_1 = k | X_0 = i) = \sum_{k=0}^{M} P_{k_j} P_{i_k}$$

Tale valore, non è altro che il valore che si troverebbe nella posizione (i, j) nella matrice data dal prodotto della matrice originale P con se stessa :

$$\begin{bmatrix} P_{0_0} & P_{0_1} & P_{0_2} & \dots & P_{0_M} \\ P_{1_0} & P_{1_1} & P_{1_2} & \dots & P_{1_M} \\ P_{2_0} & P_{2_1} & P_{2_2} & \dots & P_{2_M} \\ \vdots & \vdots & \vdots & & \vdots \\ P_{M_0} & P_{M_1} & P_{M_2} & \dots & P_{M_M} \end{bmatrix} \cdot \begin{bmatrix} P_{0_0} & P_{0_1} & P_{0_2} & \dots & P_{0_M} \\ P_{1_0} & P_{1_1} & P_{1_2} & \dots & P_{1_M} \\ P_{2_0} & P_{2_1} & P_{2_2} & \dots & P_{2_M} \\ \vdots & \vdots & \vdots & & \vdots \\ P_{M_0} & P_{M_1} & P_{M_2} & \dots & P_{M_M} \end{bmatrix} = P^{(2)} \implies P_{i_j}^{(2)} = \sum_{k=0}^{M} P_{k_j} P_{i_k}$$

Possiamo quindi *generalizzare* tale concetto, dicendo che la probabilità che si passi dallo stato i allo stato j in n passi, è il valore nella posizione (i, j) nella matrice $P^{(n)}$, ossia il risultato del prodotto della matrice P con se stessa n volte.

Esempio : Si consideri il seguente sistema a due stati : Voglio calcolare la probabilità che



$$\text{Stati} = \{0, 1\}$$

$$\mu(0) = 1 \qquad \mu(1) = 0$$

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \end{bmatrix}$$

all'istante temporale 2, ci si trovi nello stato 1, considerando che si inizia nello stato 0. Modellizzo il problema, impostando a zero la probabilità che ci si trovi nello stato 1 all'inizio. Avrò che :

$$\mathbb{P}(X_2 = 1 | X_0 = 0) = P_{1_0}^{(2)} = P_{1_0} P_{0_0} + P_{1_1} P_{1_0} = 0.5 \cdot 0.5 + 1 \cdot 0.5 = \frac{3}{4}$$

Possiamo generalizzare anche lo stato iniziale, calcolandoci la probabilità che ad un certo istante temporale n, ci si trovi in uno stato j, bisogna considerare la somma di tutte le possibili "evoluzioni" della catena, che finiscono con il trovarsi nello stato j al tempo n.

$$\sum_{k=0}^{M} \mu(k) \cdot P_{j_k}^{(n)}$$

Possiamo rappresentare in maniera naturale, un *vettore* di M coordinate contenete le probabilità che in n istanti temporali il sistema si trovi in un determinato stato fra gli $\{0, 1 \dots, M\}$ disponibili, come prima cosa definiamo il vettore :

$$\mu = \begin{bmatrix} \mu(0) & \mu(1) & \mu(2) & \dots & \mu(M) \end{bmatrix}$$

Contenente le probabilità che ha il sistema di iniziare in ognuno degli stati, consideriamo poi la matrice delle probabilità P moltiplicata per se stessa n volte.

$$P^{(n)} = \begin{bmatrix} P_{0_0} & P_{0_1} & \dots & P_{0_M} \\ P_{1_0} & P_{1_1} & \dots & P_{1_M} \\ \vdots & \vdots & & \vdots \\ P_{M_0} & P_{M_1} & \dots & P_{M_M} \end{bmatrix} \cdot \begin{bmatrix} P_{0_0} & P_{0_1} & \dots & P_{0_M} \\ P_{1_0} & P_{1_1} & \dots & P_{1_M} \\ \vdots & \vdots & & \vdots \\ P_{M_0} & P_{M_1} & \dots & P_{M_M} \end{bmatrix} \dots \text{ n volte} \cdot \begin{bmatrix} P_{0_0} & P_{0_1} & \dots & P_{0_M} \\ P_{1_0} & P_{1_1} & \dots & P_{1_M} \\ \vdots & \vdots & & \vdots \\ P_{M_0} & P_{M_1} & \dots & P_{M_M} \end{bmatrix}$$

Considereremo il vettore risultante dal prodotto riga per colonna :

$$\mu_n = \mu \cdot P^{(n)}$$

Tale vettore, alla coordinata i, contiene la probabilità che ha il sistema di trovarsi nello stato i all'n-esimo istante temporale.

Consideriamo *l'esempio* del sistema precedente a due stati, considero la matrice $P^{(2)}$, dove la posizione i, j indica la probabilità che ha il sistema di passare di passare dallo stato i allo stato j in 2 istanti :

$$P^{(2)} = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/4 & 3/4 \\ 0 & 1 \end{bmatrix}$$

Questa volta, considero :

$$\begin{cases} \mu(0) = \alpha \\ \mu(1) = \beta \end{cases} \implies \mu = \begin{bmatrix} \alpha & \beta \end{bmatrix}$$

Ottengo il vettore contenete le probabilità che all'istante n il sistema si trovi in uno dei due stati :

$$\mu_n = \begin{bmatrix} \alpha & \beta \end{bmatrix} \cdot \begin{bmatrix} 1/4 & 3/4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/4 \cdot \alpha + 3/4 \cdot \beta \\ 0 \cdot \alpha + 1 \cdot \beta \end{bmatrix} = \begin{bmatrix} \frac{\alpha}{4} + \beta\frac{3}{4} \\ \beta \end{bmatrix}$$

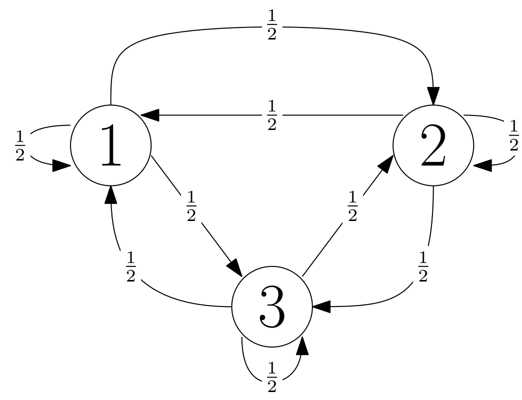
Ho quindi che, la probabilità di trovarsi nello stato 0 al secondo istante temporale è di $\frac{\alpha}{4} + \beta\frac{3}{4}$, la probabilità di trovarsi nello stato 1 è β.

5.1.2 Probabilità Stazionaria

Abbiamo definito $\mu_n = \mu \cdot P^{(n)}$ come il vettore che alla coordinata i, contiene la probabilità che ha il sistema di trovarsi nello stato i all'n-esimo istante temporale.

Ci chiediamo se esista un certo μ per la quale vale che $\mu \cdot P^{(n)} = \mu$, ad esempio, l'entrata i di μ è la probabilità che ha il sistema di trovarsi nello stato i al tempo n, dopo averlo moltiplicato a $P^{(n)}$, tale μ rimane *invariata*, si dice che μ è una probabilità **stazionaria (o invariante)**, è un *autovettore* di P con *autovalore* 1, e per convenzione lo denoteremo π.

Prendiamo in esempio il seguente grafo : Risulta ovvio che ovunque ci si trovi, la probabilità



di raggiungere un qualsiasi nodo è identica, il grafo è omogeneo e la probabilità è stazionaria. Consideriamo un esempio più generale, vi è un sistema con 2 stati, la matrice delle probabilità è $\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$, vogliamo capire se esiste un vettore di probabilità stazionario π. Dobbiamo quindi trovare π tale che $\pi \cdot P = \pi$, abbiamo $\pi = \begin{bmatrix} p & q \end{bmatrix}$, risolviamo quindi il sistema lineare :

$$\begin{bmatrix} p & q \end{bmatrix} = \begin{bmatrix} p & q \end{bmatrix} \cdot \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} = \begin{bmatrix} p & q \end{bmatrix} = \begin{bmatrix} p - p\alpha + q\beta & p\alpha + q - q\beta \end{bmatrix}$$

$$\begin{cases} p = p - p\alpha + q\beta \\ q = p\alpha + q - q\beta \\ p + q = 1 \end{cases}$$

La matrice delle probabilità è *omogenea*, ci renderemo presto conto che se il sistema ammette soluzione, essa è unica, de facto, data una certa condizione, una catena di Markov ha un *unico* vettore probabilità stazionario π, a meno di casi "degeneri", si avrà che :

$$\lim_{n\to+\infty} \mu_n = \lim_{n\to+\infty} \mu \cdot P^{(n)} = \pi$$

Definizione : Una matrice delle transizioni P è detta **ergodica** se :

$$\exists n \in \mathbb{N} | P_{i_j}^{(n)} > 0 \quad \forall i, j \in \{0, 1 \ldots, M\}$$

Euristicamente, ciò indica che è possibile andare da un qualsiasi stato/nodo ad un altro in un numero finito di passi con probabilità strettamente positiva, dato che ogni entrata della matrice ha un numero strettamente positivo.

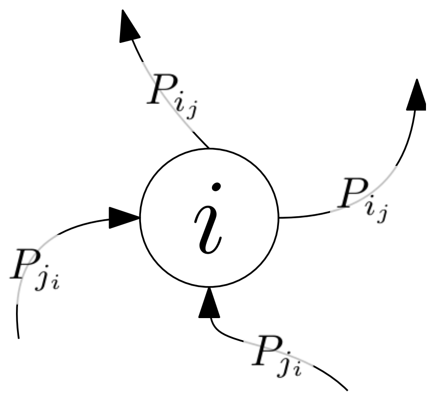
Teorema (Ergodico) : Se P è una matrice delle transizioni *ergodica*, allora :

- Esiste un unico vettore delle probabilità stazionario π.
- $\lim_{n\to+\infty} \mu \cdot P^{(n)} = \pi$.

Quindi per trovare la probabilità stazionaria bisogna risolvere il sistema $\pi = \pi \cdot P$, ma ciò può risultare estremamente inefficente quando il numero degli stati/nodi diventa alto, vi è un caso particolare nella quale è possibile trovare facilmente una soluzione esplicita.

5.1.3 Bilancio Dettagliato

Sia $V = \{0, 1 \ldots, M\}$ l'insieme degli stati/nodi della catena di Markov e P la matrice delle transizioni, sappiamo che se esiste un vettore di probabilità stazionario π, esso è unico. Considero il grafo degli stati e fisso l'attenzione su un punto $i \in V$. Il nodo i ha degli archi



uscenti e degli archi entrati, che lo collegano ad altri generici nodi j. Avremo quindi che la somma di tutte le probabilità "uscenti" dal nodo i è $\sum_{j \in V} P_{i_j}$ e che la somma di tutte le probabilità "entranti" nel nodo i è $\sum_{j \in V} P_{j_i}$, ho un certo vettore di probabilità π, e definisco :

$$\textbf{Flusso entrante in } i \quad := \sum_{j \in V} \pi_i P_{i_j} \tag{108}$$

$$\textbf{Flusso uscente da } i \quad := \sum_{j \in V} \pi_j P_{j_i} \tag{109}$$

Si impone quindi che gli archi siano simmetrici, si dice che il vettore π **soddisfa il bilancio dettagliato** se ogni nodo ha flusso entrante *identico* al flusso uscente :

$$\forall i \in V \quad \sum_{j \in V} \pi_i P_{i_j} = \sum_{j \in V} \pi_j P_{j_i}$$

Ne consegue anche che :

$$\forall i \in V \quad \pi_i = \sum_{j \in V} \pi_i P_{i_j} = \sum_{j \in V} \pi_j P_{j_i}$$

Se il vettore π soddisfa il bilancio dettagliato, allora π è il vettore di probabilità stazionario. Non è però detto il contrario.

5.1.4 Montecarlo Markov Chain