Università degli Studi di Bologna

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Matematica

Tesi di Laurea Triennale

Teorema del Limite Centrale e Legge del Logaritmo Iterato

Candidato:
Enrico Ciavatti
Matricola 655675

Relatore:
Ch.mo Prof. Nicola Arcozzi

Anno Accademico 2017-2018

Indice

Introduzione

Questo testo è volto a descrivere come si possono analizzare alcuni processi stocastici da punti di vista differenti, ossevandone, mediante strumenti matematici, il comportamento sul lungo termine. Nel primo capitolo verranno esposti tutti i concetti base atti alla comprensione di tutto ciò di cui si parlerà più avanti, in particolare verranno fatti richiami sulla Probabilità elementare e sull' Analisi Matematica, dopodiché verranno introdotti concetti più avanzati, che normalmente non vengono affrontati nei corsi standard di un c.d.l. triennale, come il valore atteso condizionato; inoltre verrà ampliato il concetto di indipendenza in Teoria della Probabilità. Il secondo capitolo sarà totalmente incentrato sulle Martingale, ed in particolare saranno presenti alcuni esempi volti a rendere più chiaro e concreto un concetto che altrimenti potrebbe apparire come ostico ed esoterico; si vedrà in questo capitolo come le martingale rappresentino un gioco equo, ossia un gioco dove sul lungo termine il guadagno e la perdita saranno destinati ed eguagliarsi. Nel terzo capitolo si parlerà di Legge forte dei Grandi Numeri e di Teorema del Limite Centrale; verrà fornita di entrambi i risultati una dimostrazione, e poi verranno interpretati e messi a confronto; ci sarà in questo capitolo una sezione a parte sul TLC dal punto di vista delle martingale, ed in particolar modo verrà introdotto il concetto di differenza di martingala, e si vedrà come esso possa essere considerato come un modo di vedere le martingale da un punto di vista differente. In questo capitolo ci si è avvalsi delle fonti *The Strong Law of Large Numbers"* di Matthias Winkel, *"Proof of Central Limit Theorem"* di H. Krieger, e di *"The Martingale Central Limit Theorem"*, di Michael Woodroofe, i cui collegamenti si trovano nella Bibliografia. Il quarto capitolo sarà dedicato nella prima sezione alle Grandi Deviazioni, dove, data una successione di ripetizioni dello stesso fenomeno aleatorio destinata ad avvicinarsi, alla lunga (ed idealmente a raggiungere dopo un numero infinito di ripetizioni) ad una situazione di stabilità (si prenderà in esame come processo campione il processo di Bernoulli) si darà una stima della difficoltà di allontanarsi da tale situazione, e si vedrà come questa difficoltà aumenti in maniera esponenziale, all'allontanarsi progressivo dalla situazione di stabilità; tutto ciò lo osserveremo con il Teorema delle Grandi Deviazioni(TGD). Nella seconda sezione del quarto capitolo si affronterà invece la Legge del Logaritmo Iterato di Khinchin, e ne verrà fornita una dimostrazione, sempre nel caso particolare di un processo di Bernoulli. Si terminerà poi con un'interpretazione della LLI, che verrà confrontata anche con la LGN forte ed il TLC; in particolare, si vedrà come la LGN forte e la LLI descrivano i processi stocastici sui singoli cammini, mentre il TLC ne descriva il comportamento in media; verranno poi analizzati i tre Teoremi sulla base della quantità di informazioni fornite, vedendo come la LGN forte tenda a comprimere troppe informazioni ed il TLC, al contrario, a lasciarne disperdere una quantità eccessiva, mentre la LLI, ponendosi in mezzo, sia molto più precisa. In questo capitolo ci si è avvalsi delle fonti *"Large Deviations"*, *"The Moderate Deviations Result"*, *Law of the Iterated Logarithm"*, tutte dalle lezioni **Topics in Probability Theory and Stochastic Processes Steven R. Dunbar**, di Steven R. Dunbar. Un processo di Bernoulli è una successione di variabili aleatorie X_n, ognuna delle quali rappresenta un tentativo effettuato, dove abbiamo due risultati complementari: 0 ed 1, ed esiste un parametro $p \in (0, 1)$ tale che 1 abbia probabilità di uscita p e 0 abbia probabilità di uscita $1 - p$; l'esempio tipico che si usa per spiegare che cosa un processo di Bernoulli rappresenta è il lancio di una monetina, dove l'uscita di una faccia è complementare all'uscita dell'altra. Detto ciò, con il processo di Bernoulli, la LGN forte dice (sotto opportune ipotesi iniziali) che $\lim_{n \longrightarrow +\infty} \frac{S_n - np}{n} = 0$, dove $S_n = X_1 + ... + X_n$; cioè, il rapporto fra il numero di volte in cui è uscito il risultato di probabilità p ed il numero di volte in cui è uscito il risultato di probabilità $1 - p$, dopo n tentativi, è destinato ad avvicinarsi progressivamente a $\frac{p}{1-p}$, al crescere di n. Il Teorema del Limite Centrale, invece, dice che (sempre sotto opportune ipotesi), $\lim_{n \longrightarrow +\infty} P\left(\frac{S_n}{\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$; qui, si cerca di dare una stima della difficolà di rientrare in una gamma di valori che stiano tutti alla sinistra, sulla retta reale, di un certo risultato x; inoltre si può notare come il limite della probabilità non dipenda da p; questo può far capire come tutti i processi aleatori simili per tipologia al lancio di una monetina tendano ad essere tutti accomunati da un comportamento dato, quindi a diventare in un certo senso indistinguibili. Il TGD, invece, dice che $P\left(\frac{Sn}{n} \geq p + \epsilon\right) \leq e^{-nh_+(\epsilon)}$, dove $h_+(\epsilon)$ è una funzione crescente definita su $(0, 1 - p)$ e ≥ 0; in pratica, questo Teorema ci descrive come la difficoltà, in un processo stocastico, di allontanarsi dalla situazione di stabilità aumenti in maniera esponenziale. Infine c'è la LLI di Khinchin, che dice che $\limsup_{n \longrightarrow +\infty} \frac{S_n - np}{\sqrt{2np(1-p)\log\log(n)}} = 1$ e che $\liminf_{n \longrightarrow +\infty} \frac{S_n - np}{\sqrt{2np(1-p)\log\log(n)}} = -1$; in pratica, si ha una continua oscillazione dei valori relativi ai rilultati che escono ad ogni tempo n, e questo Teorema ci dice in che modo questi valori oscillino, fornendoci due estremi.

Capitolo 1

Richiami di Probabilità e di Analisi Matematica.

1.1 Preambolo.

Definizione: *Sia* $\Omega \neq \emptyset$*. Allora si chiama insieme delle parti di* Ω *la famiglia* $\mathcal{P}(\Omega) = \{A | A \subseteq \Omega\}$.

Definizione: *Sia* $\Omega \neq \emptyset$*. Allora un sottoinsieme* $\mathfrak{F}$ *di* $\mathcal{P}(\Omega)$ *si dice essere una* σ*-algebra su* Ω *se valgono le seguenti condizioni:*
(σ1)$\Omega \in \mathfrak{F}$.
(σ2)$\forall A \in \mathfrak{F},\ C(A) = \Omega \setminus A \in \mathfrak{F}$.
(σ3)$\forall (A_n)_{n\in\mathbb{N}}$*, successione in* $\mathfrak{F}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F}$.

Osservazione: Se $\mathfrak{F}$ è una σ-algebra su $\Omega \neq \emptyset$, e $A, B \in \mathfrak{F}$, allora $A \setminus B$ e $B \setminus A \in \mathfrak{F}$.

Osservazione: Sia $\Omega \neq \emptyset$ e sia $\mathfrak{F} \subseteq \mathcal{P}(\Omega)$. Allora, se $\mathfrak{F}$ gode delle proprietà (σ1) e (σ2), $\mathfrak{F}$ è una σ-algebra sse $\forall (A_n)_{n\in\mathbb{N}}$ successione in $\mathfrak{F}$, $\bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{F}$.

Osservazione: Un'intersezione di σ-algebre è ancora una σ-algebra.

Osservazione: Se $\mathfrak{F}$ è una σ-algebra su $\Omega \neq \emptyset$, e $E \in \mathcal{P}(\Omega)$ è tale che $\exists B \in \mathfrak{F} : E \cap B, E \cap C(B) \in \mathfrak{F}$, allora $E \in \mathfrak{F}$.

Definizione: *Sia* $\mathfrak{F}$ *una* σ*-algebra su un insieme* $\Omega \neq \emptyset$*. Allora* $\mathfrak{G} \subseteq \mathfrak{F}$ *si dice essere una sotto*σ*-algebra di* $\mathfrak{F}$ *se è ancora una* σ*-algebra. In questo testo, useremo la scrittura* $\mathfrak{G} \leq \mathfrak{F}$.

Definizione: *Sia* $\Omega \neq \emptyset$ *e sia* $H \subseteq \mathcal{P}(\Omega)$*. Allora si chiama* σ*-algebra generata da* H *la più piccola* σ*-algebra su* Ω *contenente* H *o, equivalentemente, l'intersezione di tutte le* σ*-algebre contenenti* H*. La indicheremo con* $\sigma(H)$.

Definizione: *Su* $\mathbb{R}$*, le famiglie di insiemi* $\{[a,b] \,|a < b\}$, $\{[a,b) \,|a < b\}$, $\{(a,b] \,|a < b\}$, $\{(a,b) \,|a < b\}$, $\{(-\infty,b] \,|b \in \mathbb{R}\}$, $\{[a,+\infty) \,|a \in \mathbb{R}\}$, $\{(-\infty,b) \,|b \in \mathbb{R}\}$ *e* $\{(a,+\infty) \,|a \in \mathbb{R}\}$

generano tutte la stessa σ-algebra. Tale σ-algebra si chiama σ-algebra di Borel su $\mathbb{R}$, e la indicheremo con $\mathcal{B}(\mathbb{R})$.

Osservazione: Più avanti, ci tornerà utile anche utilizzare la σ-algebra di Borel su $[0,1]$, ossia $\mathcal{B}([0,1])=\sigma(\{[a,b]\,|0 \leq a < b \leq 1\})$.

Definizione: *Sia $\Omega \neq \emptyset$ e sia $\mathfrak{F}$ una σ-algebra su di esso. Allora una funzione P:$\mathfrak{F} \longrightarrow [0,1]$ si dice essere una misura di probabilità su Ω o, più semplicemente, probabilità, se:*
(1)$P(\Omega)$=1.
(2)$\forall A, B \in \mathfrak{F}$, se $A \subseteq B$, allora $P(A) \leq P(B)$.
(3)$\forall (A_n)_{n\in\mathbb{N}}$ successione in $\mathfrak{F}$ vale che $P(\bigcup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} P(A_n)$ e vale l'uguaglianza sse $\forall h, k \in \mathbb{N}$, $A_h \cap A_k = \emptyset$ qualora $h \neq k$.
Chiameremo eventi gli elementi di $\mathfrak{F}$ e chiameremo spazio di probabilità la tripla $(\Omega, \mathfrak{F}, P)$.

Osservazione:
◇ Se $A, B \in \mathfrak{F}$, e $A \subseteq B$, allora $P(B \setminus A) = P(B) - P(A)$.
◇ Se $A, B \in \mathfrak{F}$, e $P(A \cap B) = 0$, allora $P(A \cup B) = P(A) + P(B)$.

Uno spazio di probabilità si dice *completo* se dato un qualunque evento di probabilità nulla, i suoi sotto insiemi sono ancora elementi della σ-algebra che stiamo prendendo in considerazione. Ogni spazio Ω può essere completato estendendo la sua σ-algebra con elementi di $\mathcal{P}(\Omega)$ compresi fra due eventi che hanno la stessa probabilità.

Definizione: *Due eventi si dicono indipendenti se la probabilità della loro intersezione è uguale al prodotto delle loro probabilità.*

Definizione: *Dati due eventi A e B, si chiama probabilità di B condizionata ad A, la quantità: $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ se $P(A) \neq 0$, $P(B|A) = 0$ se $P(A) = 0$.*

Osservazione: Se A e B sono indipendenti e $P(A) \neq 0$, allora $P(B|A) = P(B)$.

Definizione: *Chiamiamo $\mathbb{R}$ esteso l'insieme $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$.*

Si può definire $\mathcal{B}(\bar{\mathbb{R}})$ in maniera analoga a come è stato definito $\mathcal{B}(\mathbb{R})$, ponendo $\mathcal{B}(\bar{\mathbb{R}}) = \sigma(\big\{[-\infty, b]|b \in \bar{\mathbb{R}}\big\}) = \sigma(\big\{[a, +\infty]|a \in \bar{\mathbb{R}}\big\})$.

Definizione: *Sia dato uno spazio di probabilità non vuoto $(\Omega, \mathfrak{F}, P)$. Allora diremo che una funzione $f : \Omega \longrightarrow \bar{\mathbb{R}}$ è $\mathfrak{F}$-misurabile se $\forall B \in \mathcal{B}(\mathbb{R})$, $f^{-1}(B) \in \mathfrak{F}$.*

Definizione: *Dato uno spazio di probrabilità non vuoto, chiamiamo variabile aleatoria su Ω una qualunque funzione $X : \Omega \longrightarrow \bar{\mathbb{R}}$ che sia $\mathfrak{F}$-misurabile.*

Definizione: *Diciamo che due variabili aleatorie X e Y sono indipendenti se $\forall A, B \in \mathfrak{F}$, $P(X \in A \wedge Y \in B) = P(X \in A) \cdot P(Y \in B)$.*

Definizione: *Data una variabile aleatoria $X : \Omega \longrightarrow \bar{\mathbb{R}}$, chiamiamo distri-*

buzione di probabilità di X l'applicazione $D_X : H \in \mathcal{B}(\bar{\mathbb{R}}) \longmapsto P(X \in A)$. Scriveremo $X \sim D_X$.

1.2 Integrazione su uno spazio di misura.

Definizione: *Sia $\Omega \neq \emptyset$, allora chiamiamo misura su Ω una funzione:*

$$m : \mathcal{P}(\Omega) \longrightarrow [0, +\infty]$$

tale che:
(i) $m(\emptyset) = 0$.
(ii) $m(A) \leq m(B)$ se $A \subseteq B \subseteq \Omega$ (m è monotòna).
(iii) $m(\bigcup_{k\in\mathbb{N}} A_k) \leq \sum_{k\in\mathbb{N}} m(A_k)$ (m è numerabilmente sub-additiva).

Osservazione: Ogni misura su Ω è finitamente sub-additiva. Più esplicitamente:

$$m(\cup_{k=1}^{p} A_k) \leq \sum_{k=1}^{p} m(A_k) \forall p \geq 1.$$

Definizione: *Un sottoinsieme E di Ω si dice m-misurabile se $\forall A \subseteq \Omega$, si ha che:*

$$m(E) = m(E \cap A) + m(E \cap \mathcal{C}(A)).$$

Osservazione: Si può dimostrare che la famiglia degli insiemi m-misurabili forma una σ-algebra su Ω, e che, se $(A_k)_{k\in\mathbb{N}}$ è una successione disgiunta di insiemi m-misurabili, allora $m(\bigcup_{k\in\mathbb{N}}) = \sum_{k=1}^{+\infty} m(A_k)$.

Osservazione: Se due insiemi $A, B \subseteq \Omega$ sono m-misurabili e $A \subseteq B$, allora $m(B \setminus A) = m(B) - m(A)$

Osservazione: Se $(A_k)_{k\in\mathbb{N}}$ è una successione di sottoinsiemi m-misurabili di Ω e se $(A_k)_{k\in\mathbb{N}}$ è monotona crescente (i.e. $A_k \subseteq A_{k+1}$ $\forall k \in \mathbb{N}$), allora:

$$m(\bigcup_{k\in\mathbb{N}} A_k) = \lim_{k\longrightarrow+\infty} m(A_k).$$

Osservazione: Se $(A_k)_{k\in\mathbb{N}}$ è una successione di sottoinsiemi m-misurabili di Ω, e se $(A_k)_{k\in\mathbb{N}}$ è monotona decrescente (i.e. $A_k \supseteq A_{k+1}$ $\forall k \in \mathbb{N}$) e dove $m(A_1) < \infty$, allora:

$$m(\bigcap_{k\in\mathbb{N}} A_k) = \lim_{k\longrightarrow+\infty} m(A_k).$$

Data una certa condizione ϕ, diremo che essa vale *quasi ovunque*(qu.o.), o *quasi sicuramente*(qu.s.) su Ω se il più grande sottoinsieme m-misurabile di Ω in cui ϕ non è verificata ha misura nulla.

Adesso, studiamo l'integrazione rispetto a una misura astratta.

Definizione: *Siano dati un insieme $\Omega \neq \emptyset$ ed una misura m su Ω. Se A è un sottoinsieme m-misurabile di Ω, chiamiamo m-scomposizione di A una famiglia $\sigma = (A_k)_{k\in\mathcal{A}}$ di sottoinsiemi di A tale che:*
(i)$\mathcal{A}$ è finito o al più numerabile.
(ii)A_k è m-misurabile $\forall k \in \mathcal{A}$.
(iii)$\bigcup_{k\in\mathcal{A}} A_k = A$ e $A_h \cap A_k = \emptyset$ se $k \neq h$.

Indichiamo con $\Omega_A(m)$ la totalità delle m-scomposizioni di A.

Sia ora $f : A \longrightarrow [0, +\infty]$ una funzione non negativa sull'insieme A. Per ogni m-scomposizione $\sigma = (A_k)_{k\in\mathcal{A}}$ di A, poniamo:

$$S(f,\sigma) = \sum_{k\in\mathcal{A}} sup f_{A_k} m(A_k),$$

$$s(f,\sigma) = \sum_{k\in\mathcal{A}} inf f_{A_k} m(A_k).$$

Si chiamano integrale superiore e integrale inferiore rispetto a m, della funzione f su A, rispettivamente i numeri reali estesi

$$\int_A^{-} fdm := inf \left\{ S(f,\sigma) | \sigma \in \Sigma_A(m) \right\},$$

$$\int_{-}^{A} fdm := sup \left\{ s(f,\sigma) | \sigma \in \Omega_A(P) \right\}.$$

Si dice che f è m-integrabile su A se:

$$\int_{-}^{A} fdm = \int_A^{-} fdm$$

e chiamiamo integrale di f su A rispetto ad m, la quantità

$$\int_A fdm := \int_{-}^{A} fdm = \int_A^{-} fdm.$$

Definizione: *Sia A un sottoinsieme di Ω m-misurabile e sia $f : A \longrightarrow \bar{\mathbb{R}}$. Si dice che f è m-misurabile su A se l'insieme $\{f < c\}$ è m-misurabile $\forall c \in \mathbb{R}$.*

Osservazione: $\{f < c\}$ si può sostituire con $\{f \leq c\}$, $\{f > c\}$ o $\{f \geq c\}$.

Teorema: *Sia A un sottoinsieme di Ω m-misurabile e sia $f : A \longrightarrow [0, \infty]$. Se f è m-misurabile, allora f è m-integrabile. Viceversa, se f è m-integrabile e se*

$$\int_A fdm < +\infty,$$

allora f è m-misurabile.

Estendiamo ora la nozione di integrabilità alle funzioni a valori in $\bar{\mathbb{R}}$.

Sia $f : A \longrightarrow \bar{\mathbb{R}}$ una qualunque funzione misurabile su A. Allora, si chiamano parte positiva e parte negativa di f, rispettivamente le funzioni $f^+ = max\{0, f\}$ ed $f^- = max\{0, -f\}$.
Le funzioni f^+ ed f^- sono m-misurabili (poiché lo è f) e non negative, esse sono quindi m-integrabili. Diciamo che f è integrabile su A se almeno uno degli integrali

$$\int_A f^+ dm, \int_A f^- dm$$

è finito; in questo caso, poniamo:

$$\int_A f dm = \int_A f^+ dm - \int_A f^- dm$$

chiamando tale quantità integrale di f su A.

1.3 Spazi Lp.

Sia dato uno spazio di misura (Ω, m) e sia $1 \leq p < \infty$. Allora, chiamiamo $\mathfrak{L}^p(\Omega, m)$ l'insieme delle funzioni f m-misurabili tali per cui $\int_\Omega |f|^p dm < +\infty$. Si può dimostrare che $\mathfrak{L}^p(\Omega, m)$ forma un $\mathbb{R}$-spazio vettoriale, se si considera, presi $\lambda, \mu \in \mathbb{R}$ ed $f, g \in \mathfrak{L}^p(\Omega, m)$, la funzione $\lambda f + \mu g : x \in \Omega \longmapsto \lambda f(x) + \mu g(x)$, e si identifica con 0 la funzione identicamente nulla. Introduciamo, ora, una relazione di equivalenza su $\mathfrak{L}^p(\Omega, m)$: prese $f, g \in \mathfrak{L}^p(\Omega, m)$, diciamo che $f \simeq g$ sse $f - g = 0$ qu.o.(cioè a meno di un insieme di misura nulla); definiamo, adesso, l'insieme $L^p(\Omega, m)$ come l'insieme di tutte le classi di equivalenza di $\simeq$. Se $f \in \mathfrak{L}^p(\Omega, m)$, allora, con abuso di linguaggio, scriveremo che $f \in L^p(\Omega, m)$. Poiché la somma è compatibile con $\simeq$, anche $L^p(\Omega, m)$ diventa un $\mathbb{R}-$spazio vettoriale, identificando la classe delle funzioni identicamente nulle con 0. Poiché la probabilità è un particolare tipo di misura astratta, allora ha senso parlare di integrazione su uno spazio di probabilità. Se $(\Omega, \mathfrak{F}, P)$ è uno spazio di probabilità, allora parleremo di $\mathfrak{F}$-misurabilità, associando P ad $\mathfrak{F}$, e lo spazio $L^p(\Omega, P)$, diventerà semplicemente $L^p(\Omega)$, o ancora più semplicemente L^p.

1.4 Il valore atteso, o speranza matematica.

Definizione: *Sia $(\Omega, \mathfrak{F}, P)$ uno spazio di probabilità e sia $Y : \Omega \longrightarrow \bar{\mathbb{R}}$ una variabile aleatoria su di esso. Allora si chiama σ-algebra generata da Y l'intersezione di tutte le $\mathfrak{G} \leq \mathfrak{F}$, tali che Y sia $\mathfrak{G}$-misurabile. Indichiamo con $\mathfrak{F}_Y$ tale σ-algebra.*

Si può dimostrare che, data una variabile aleatoria $X : \Omega \longrightarrow \bar{\mathbb{R}}$, essa è $\mathfrak{F}_Y$-misurabile sse $\exists\ \varphi : \mathbb{R} \longrightarrow \mathbb{R}$ Borel-misurabile tale che $X = \varphi(Y)$.

Definizione: *Sia $(\Omega, \mathfrak{F}, P)$ uno spazio di probabilità e sia $X : \Omega \longrightarrow \bar{\mathbb{R}}$ una* *variabile aleatoria su di esso. Allora, chiamiamo valore atteso (o speranza matematica) di X, la quantità:*

$$E[X] = \sum_{x \in \bar{\mathbb{R}}} x P(X = x).$$

Definizione: *Sia $(\Omega, \mathfrak{F}, P)$ uno spazio di probabilità, sia $X : \Omega \longrightarrow \bar{\mathbb{R}}$ una variabile aleatoria e sia $A \in \mathfrak{F}$. Allora chiamiamo valore atteso di X condizionato ad A, la quantità:*

$$E[X|A] = \sum_{x \in \mathbb{R}} x P(X = x|A).$$

Osservazione: Siano $\lambda, \mu \in \mathbb{R}$ e siano X ed Y due variabili aleatorie. Allora:

$$E[\lambda X + \mu Y|A] = \lambda E[X|A] + \mu E[Y|A].$$

Osservazione: $E[X|\Omega] = E[X]$.

Come prima cosa, osserviamo che il valore atteso condizionato è un'estensione del concetto di probabilitá condizionata: infatti, se B è un evento e I_B è una v.a. tale che $P(I_B = 1|A) = P(B|A)$ e $P(I_B = 0|A) = 1 - P(B|A)$ (chiameremo I_B funzione indicatrice di B; essa sarà unica qu.s.), allora $P(B|A) = E[I_B|A]$. Cioè, possiamo ottenere una probabilità condizionata dal valore atteso condizionato di una funzione indicatrice.
Come seconda cosa, si può dimostrare che, in generale, se $(\Omega, \mathfrak{F}, P)$ è uno spazio di probabilità, $X : \Omega \longrightarrow \mathbb{R}$ è una variabile aleatoria e $A \in \mathfrak{F}$, allora vale il seguente risultato:

$$E[X|A] = \frac{1}{P(A)} \int_A X dP.$$

Da questo risultato, si può dedurre che il valore atteso condizionato non è altro che la media dei valori che una variabile aleatoria assume su un determinato insieme.

Definizione: *Sia $(\Omega, \mathfrak{F}, P)$ uno spazio di probabilitá e siano $X : \Omega \longrightarrow \mathbb{R}$ ed $Y : \Omega \longrightarrow \mathbb{R}$ due variabili aleatorie, con Y discreta(ossia con immagine finita). Allora si chiama valore atteso di X condizionato ad Y, la funzione:*

$$E[X|Y](\omega) = \sum_{y \in \mathbb{R}} E[X|Y = y] I_{Y=y}(\omega) \forall \omega \in \Omega.$$

Osservazione: Tale funzione, è una variabile aleatoria su Ω.

Teorema: *Sia $(\Omega, \mathfrak{F}, P)$ uno spazio di probabilità e siano $X : \Omega \longrightarrow \bar{\mathbb{R}}$ e $Y : \Omega \longrightarrow \bar{\mathbb{R}}$ due variabili aleatorie, con Y discreta. Allora:*

(i)$E[X|Y]$ è $\mathfrak{F}_Y$-misurabile.

(ii)$\int_\Lambda E[X|Y]dP = \int_\Lambda XdP$ $\forall \Lambda \in \mathfrak{F}_Y$.

Adesso, estendiamo il concetto di $E[X|Y]$:

Definizione: *Sia $(\Omega, \mathfrak{F}, P)$ uno spazio di probabilità e sia $\mathfrak{G} \leq \mathfrak{F}$. Se X è una v.a.i.(variabile aleatoria integrabile), allora chiamiamo valore atteso di X condizionato a $\mathfrak{G}$, una qualunque variabile aleatoria Z che goda delle proprietà seguenti:*

(CE1) Z è $\mathfrak{G}$-misurabile

(CE2) $\int_\Lambda ZdP = \int_\Lambda XdP$ $\forall \Lambda \in \mathfrak{G}$.

Denotiamo Z con $E[X|\mathfrak{G}]$.

Osservazione: È implicito in (CE1) che Z debba essere integrabile.

Vi sono alcuni quesiti che sorgono spontanei:
(1)Qual è il ruolo delle σ-algebre in tutto ciò? La risposta è che una σ-algebra rappresenta degli eventi: ad esempio, se $\mathfrak{G}$ è generata da una v.a. Y, allora $\mathfrak{G}$ conterrà tutti gli eventi del tipo $(Y \in B)$ con $B \in \mathcal{B}(\bar{\mathbb{R}})$.
(2)Perché il valore atteso condizionato dovrebbe essere una variabile aleatoria? Supponiamo che $\mathfrak{G}$ sia generata da Y: allora, ci si aspetta che $E[X|\mathfrak{G}] = E[X|Y]$, e quest'ultima è una funzione di Y; se Y è una variabile aleatoria, allora lo sarà anche una qualunque sua funzione.
(3)Come interviene il concetto di misurabilità? Se $\mathfrak{G} = \mathfrak{F}_Y$, allora il valore atteso dovrebbe essere una funzione di Y, come appena visto. Ora, se Z è $\mathfrak{F}_Y$-misurabile, allora è funzione di Y.
(4)Che cosa significa (CE2)? Si tratta di una generalizzazione del fatto che:

$$\int_{Y=y} E[X|Y]dP = \int_{Y=y} XdP.$$

Cioè, il valore atteso condizionato è una media di X su un certo insieme. Poiché (CE2) è una generalizzazione dell'ultima uguaglianza vista, allora si può dire che $E[X|\mathfrak{G}]$ sia, in un certo senso, una media che X assume sugli insiemi di $\mathfrak{G}$.

Qui di seguito, una serie di risultati riguardanti la speranza matematica condizionata, assumendo in tutti i casi di avere uno spazio di probabilità $(\Omega, \mathfrak{F}, P)$ e $\mathfrak{G} \leq \mathfrak{F}$.

Proposizione: *Se Z e Z' sono due variabili aleatorie soddisfacenti (CE1) e (CE2), allora $Z = Z'$ quasi sicuramente.*

Proposizione: *Sia X una v.a.i. e sia $\mathfrak{G} \leq \mathfrak{F}$ la σ-algebra generata da una partizione $(\Lambda_i)_{i\in\mathcal{I}}$ di Ω. Allora, vale che:*

$$E[X|\mathfrak{G}] = \sum_{i\in\mathcal{I}} \frac{E[XI_{\Lambda_i}]}{P(\Lambda_i)} I_{\Lambda_i}.$$

In particolare, $E[X|Y] = E[X|\mathfrak{F}_Y]$.

Osservazione: La funzione indicatrice di un insieme B la considereremo come quella funzione unica qu.s. tale da assumere valore 1 su B e valore 0 nei punti che non stanno in B.

Teorema: *Siano* X *ed* Y *due v.a.i., e siano* $a, b \in \mathbb{R}$. *Dunque:*
(i)$E[E[X|\mathfrak{G}]] = E[X]$.
(ii)Se $\mathfrak{G} = \{\Omega; \emptyset\}$, *allora* $E[X|\mathfrak{G}] = E[X]$ *qu.o.*
(iii)Se X *è* $\mathfrak{G}$*-misurabile, allora* $E[X|G] = X$ *qu.o.*
(iv)$E[aX + bY|\mathfrak{G}] = aE[X|\mathfrak{G}] + bE[Y|\mathfrak{G}]$.
(v)Se $X \geq 0$ *qu.s., allora* $E[X|\mathfrak{G}] \geq 0$ *qu.s.*
(vi)Se $X \leq Y$ *qu.s., allora* $E[X|\mathfrak{G}] \leq E[Y|\mathfrak{G}]$ *qu.s.*
(vii) $|E[X|\mathfrak{G}] \leq E[|X||\mathfrak{G}]$ *qu.o.*
(viii) Supponiamo che Y *sia* $\mathfrak{G}$*-misurabile e che* XY *sia integrabile. Allora si ha, quasi ovunque, che:*

$$E[XY|\mathfrak{G}] = YE[X|\mathfrak{G}].$$

(ix) Se X_n *e* X *sono integrabili, e* $X_n \nearrow X$ *oppure* $X_n \searrow X$, *allora:*

$$E[X_n|\mathfrak{G}] \longrightarrow E[X|\mathfrak{G}].$$

Teorema: *Se* X *è una v.a.i. e* $\mathfrak{G}_1 \leq \mathfrak{G}_2$, *allora:*

$$E[E[X|\mathfrak{G}_1]|\mathfrak{G}_2] = E[E[X|\mathfrak{G}_2]|\mathfrak{G}_1] = E[X|\mathfrak{G}_1].$$

Corollario: *Sia* X *una v.a.i. e siano* $\mathfrak{G}_1 \leq \mathfrak{G}_2$. *Allora* $E[X|\mathfrak{G}_2] = E[X|\mathfrak{G}_1]$ *sse* $E[X|\mathfrak{G}_2]$ *è* $\mathfrak{G}_1$*-misurabile.*

Definizione: *Una funzione* $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$ *si dice convessa se* $\forall a, b \in \mathbb{R}$ *e* $\forall \lambda \in [0, 1]$, $\Phi(\lambda a + (1 - \lambda)b) \leq \lambda\Phi(a) + (1 - \lambda)\Phi(b)$.

Ora esplichiamo un importante risultato, che prende il nome di *Disuguaglianza di Jensen*:

Teorema: *Sia* $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$ *una funzione convessa e sia* X *una variabile aleatoria. Allora, se* X *e* $\Phi(X)$ *sono entrambe integrabili, vale che:*

$$\Phi(E[X]) \leq E[\Phi(X)].$$

Teorema: *Sia* X *una v.a. e sia* Φ *una funzione convessa; supponiamo che* X *e* $\Phi(X)$ *siano entrambe integrabili, e che* $\mathfrak{G} \leq \mathfrak{F}$. *Allora, qu.o., vale che:*

$$\Phi(E[X|\mathfrak{G}]) \leq E[\Phi(X)|\mathfrak{G}].$$

L'ultimo Teorema visto, è una generalizzazione della Disuguaglianza di Jensen.

Generalizziamo, ora, il concetto di *indipendenza*, con una serie di definizioni ed un risultato:

Definizione: *Due σ-algebre $\mathfrak{G}$ ed $\mathfrak{H}$ si dicono indipendenti se $\forall \Lambda \in \mathfrak{G} \wedge \forall \Gamma \in \mathfrak{H}$, vale che $P(\Lambda \cap \Gamma) = P(\Lambda)P(\Gamma)$.*

Definizione: *Una famiglia di σ-algebre $(\mathfrak{G}_i)_{i \in \mathcal{I}}$, con $\mathcal{I}$ finito o al più numerabile, si dice indipendente se $\forall i, k \in \mathcal{I}$, con $i \neq k$, $\mathfrak{G}_i$ e $\mathfrak{G}_k$ sono indipendenti.*

Definizione: *Una famiglia finita di σ-algebre $\mathfrak{G}_1, ..., \mathfrak{G}_n$ si dice indipendente se $\forall \Lambda_i \in \mathfrak{G}_\mathfrak{i}$, con $i = 1, ..., n$, si ha che:*

$$P(\cap_{i=1}^n \Lambda_i) = \prod_{i=1}^n P(\Lambda_i).$$

Definizione: *Una famiglia infinita di sigma-algebre si dice indipendente se lo è ogni sua sottofamiglia finita.*

Proposizione: *Sia X una v.a. e sia $\mathfrak{G}$ una σ-algebra. Supponiamo che X e $\mathfrak{G}$ siano indipendenti (vale a dire: $\mathfrak{F}_X$ e $\mathfrak{G}$ sono indipendenti). Allora $E[X|\mathfrak{G}] = E[X]$ quasi ovunque.*

Osservazione: Se X è indipendente da $\mathfrak{G}$, allora, $\forall \Lambda \in \mathfrak{G}$, si ha che:

$$E[X|\Lambda] = \sum_{x \in \bar{\mathbb{R}}} xP(X = x|\Lambda) = \sum_{x \in \bar{\mathbb{R}}} xP(X = x) = E[X]$$

(X è $\mathfrak{F}$-misurabile, pertanto $(X = x) \in \mathfrak{F}$).

Capitolo 2

Martingale.

Le Martingale rappresentano un modello matematico che descrive un gioco aleatorio equo, ossia, idealmente, un gioco in cui, se $p \in (0,1)$ è la probabilità di vincere ad ogni signola giocata, allora in una singola giocata, chiamato G il guadagno che si effettuerebbe in caso di vincita, e chiamata P la perdita che si subirebbe in caso di mancata vincita, $G \cdot p = P \cdot (1-p)$, cioè $\frac{G}{P} = \frac{1}{p} - 1$.

Definizione: *Una filtrazione su uno spazio di probabilità* $(\Omega, \mathfrak{F}, P)$ *è una successione* $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ *di sottosigma-algebre di* $\mathfrak{F}$ *tali che* $\mathfrak{F}_n \subseteq \mathfrak{F}_{n+1}$, $\forall n \in \mathbb{N}$.

Qui, $\mathfrak{F}_\mathfrak{n}$ può essere pensata come l'informazione disponibile al tempo n: se $A \in \mathfrak{F}_\mathfrak{n}$, allora per ogni $m \geq n$, possiamo decidere se $\omega \in A$.

Chiameremo *spazio di probabilità filtrato* la quadrupla $(\Omega, \mathfrak{F}, (\mathfrak{F}_n)_{n\in\mathbb{N}}, P)$.

Definizione: *Un processo stocastico è una qualunque famiglia di v.a. definite sullo stesso spazio di probabilità.*

Definizione: *Un processo stocastico* X=$\{X_n, n = 0, 1, 2, 3...\}$ *si dice adattato alla filtrazione* $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ *se,* $\forall n = 0, 1, 2...$, X_n *è* $\mathfrak{F}_n$*-misurabile.*

Definizione: *Un processo* X=$(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ *si dice essere una martingala se* $\forall n \in \mathbb{N}$, *si ha che:*
(i)$(\mathfrak{F}_n)_{n\in\mathbb{N}}$ *è una filtrazione ed* X *è adattato a* $(\mathfrak{F}_n)_{n\in\mathbb{N}}$.
(ii)X_n *è integrabile.*
(iii)$E[X_{n+1}|\mathfrak{F}_n] = X_n$.

Osservazione: Verificata la (i), poi nella (iii) basta verificare (CE2).

Qui di seguito verranno proposti alcuni esempi di martingala.

Esempio 1: Ω=[0,1]; $\mathfrak{F} = \mathcal{B}([0,1])$; prendiamo la misura di Lebesgue su [0,1] come misura di probabilità. Ora, $\forall n = 0, 1, 2, ...$, costruiamo la seguente famiglia di σ-algebre:
$\mathfrak{F}_0 = \sigma(\{(0,1]\})$

$\forall n \geq 1$, $\mathfrak{F}_n = \sigma(\{(\frac{j-1}{2^n}, \frac{j}{2^n}]\}_{j=1}^{2^n})$
Sia ora $f \in L^1([0,1])$, e costruiamo una successione $f_n : [0,1] \longrightarrow \mathbb{R}$ in questo modo:

$$f_0(x) \equiv \int_0^1 f(x)dx = f_0 = x_0^0$$

mentre, $\forall n \geq 1$ e $\forall j = 1, ..., 2^n$, poniamo:

$$x_j^n = 2^n \int_{\frac{j-1}{2^n}}^{\frac{j}{2^n}} f(x)dx$$

e definiamo:

$$f_n(x) = \sum_{j=1}^{2^n} x_j^n X_{I_j^n}(x),$$

dove I_j^n=$[\frac{j-1}{2^n}, \frac{j}{2_n}] \setminus \{\frac{j-1}{2^n}\}$ e $X_{I_j^n}$ è la sua funzione indicatrice.
Dimostriamo, adesso, che il processo stocastico $(f_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ è una martingala:
(i)Osserviamo innanzitutto che, in generale, se $\Omega \neq \emptyset$ e $H_1, H_2 \subseteq \mathcal{P}(\Omega)$ sono tali che $\forall x \in H_1, \exists y, z \in H_2 : x = y \cup z$, allora $\sigma(H_1) \subseteq \sigma(H_2)$; pertanto, alla luce di questa osservazione, risulta che $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ è una filtrazione.
Sia ora $n \in \mathbb{N}$, allora f_n è una funzione semplice, per come è stata definita, e $\forall c \in \mathbb{R}$, si ha che $\{f \leq c\} = \cup_{j:f_n|_{I_j^n} \leq c} \in \mathfrak{F}_n$, dunque, f_n è $\mathfrak{F}_n$-misurabile.
(ii)Se $n \in \mathbb{N}$, allora:

$$\int_0^1 f_n(x)dx = \int_0^1 \sum_{j=1}^{2^n} x_j^n X_{I_j^n}(x)dx = \sum_{j=1}^{2^n} \int_0^1 X_{I_j^n}(x)dx = \sum_{j=1}^{2^n} x_j^n \int_{\frac{j-1}{2^n}}^{\frac{j}{2^n}} dx =$$

$$= \sum_{j=1}^{2^n} x_j^n \frac{1}{2^n} = \sum_{j=1}^{2^n} \frac{2^n}{2^n} \int_{\frac{j-1}{2^n}}^{\frac{j}{2^n}} f(x)dx = \int_0^1 f(x)dx = f_0.$$

Quindi f_n è integrabile.
(iii)Sia $n \in \mathbb{N}$, sia $\Lambda \in \mathfrak{F}_n$, e verifichiamo (CE2):
Innanzitutto, possiamo scrivere Λ come un'unione disgiunta: $\Lambda = \cup_{k\in\mathcal{K}} I_k^n$ con $|\mathcal{K}| \leq n$. Allora:

$$\int_\Lambda f_n(x)dx = \int_{\bigcup_{k\in\mathcal{K}} I_k^n} f_n(x)dx = \sum_{k\in\mathcal{K}} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} f_n(x)dx = \sum_{k\in\mathcal{K}} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} f(x)dx =$$

$$\sum_{k\in\mathcal{K}} (\int_{\frac{k-1}{2^n}}^{\frac{2k-1}{2^{n+1}}} f(x)dx + \int_{\frac{2k-1}{2^{n+1}}}^{\frac{k}{2^n}} f(x)dx) = \sum_{k\in\mathcal{K}} (\int_{\frac{k-1}{2^n}}^{\frac{2k-1}{2^{n+1}}} f_{n+1}(x)dx + \int_{\frac{2k-1}{2^{n+1}}}^{\frac{k}{2^n}} f_{n+1}(x)dx)$$

$$= \sum_{k\in\mathcal{K}} \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} = \int_{\bigcup_{k\in\mathcal{K}} I_k^n} f_{n+1}(x)dx = \int_\Lambda f_{n+1}(x)dx.$$

Esempio 2: Sia data una variabile aleatoria h che ogni qualvolta venga chiamata restituisce uno di due valori $D_1 \neq D_2$, in maniera indipendente dalla precedente, dove D_1 ha probabilità di uscita $\psi \in (0,1)$ e D_2 ha probabilità di uscita $1-\psi$. Sia Ω l'insieme delle successioni ad elementi in $\{D_1, D_2\}$, che identificheremo con $\{D_1, D_2\}^\infty$. Adesso, fissati un $n \in \mathbb{N}$ ed un $k \leq n$, definiamo $\Omega^{n,k}$ come l'insieme degli elementi di Ω che hanno k volte D_1 ed $n-k$ volte D_2 nelle prime n posizioni. Definiamo ora una misura di probabilità $P : \mathcal{P}(\Omega) \longrightarrow [0,1]$, tale da mandare l'insieme vuoto e i singoletti in 0, e tale che $\Omega^{n,k} \longmapsto \psi^k(1-\psi)^{n-k}\binom{n}{k}$. Questa definizione ha senso: infatti $\Omega = \cup_{k=0}^n \Omega^{n,k}$ unione disgiunta per ogni n naturale; dunque, $P(\Omega) = P(\cup_{k=0}^n \Omega^{n,k}) = \sum_{k=0}^n P(\Omega^{n,k}) = \sum_{k=0}^n \psi^k(1-\psi)^{n-k}\binom{n}{k} = 1$. Quindi, Ω, munito del suo insieme delle parti e della misura di probabilità appena definita, si può considerare come una rappresentazione astratta di tutte le sequenze infinite di chiamate della h. Detto ciò, costruiamo la seguente famiglia di σ-algebre:

$$\mathfrak{F}_0 = \{\Omega, \emptyset\},$$

$$\mathfrak{F}_{\mathfrak{n}+\mathfrak{1}} = \sigma(\{\Omega^{n+1,k} \cap A\})_{0 \leq k \leq n+1 \wedge A \in \mathfrak{F}_n} \forall n \geq 0.$$

Allora si tratta di una filtrazione: infatti, se $B \in \mathfrak{F}_n$, con $n \geq 0$, allora $B = \cup_{k=0}^{n+1} B \cap \Omega^{n+1,k} \in \mathfrak{F}_{n+1}$.
Adesso, $\forall \omega \in \Omega$, e $\forall i \geq 1$, poniamo: $\omega'_i = 1$ se $\omega_i = \mathrm{D}_1$, mentre $\omega'_i = -\frac{\psi}{1-\psi}$, se $\omega_i = \mathrm{D}_2$, e consideriamo il processo stocastico $Z_n : \omega \in \Omega \longmapsto \sum_{i=1}^n \omega'_i$, $\forall n \geq 1$.
Dimostriamo che il processo $(Z_n, \mathfrak{F}_n)_{n \geq 1}$ è una martingala:

(i)Sia $c \in \mathbb{R}$; allora $\{Z_n \leq c\} = \cup_{0 \leq k \leq p(n-c)+c} \Omega^{n,k} \in \mathfrak{F}_n$.

(ii)$\forall n \geq 1$, $\int_\Omega Z_n dP = E[Z_n] = \sum_{z \in \mathbb{R}} zP(Z_n = x)$

$= \sum_{k=0}^n \frac{k-\psi n}{1-\psi} \psi^k (1-\psi)^{n-k}\binom{n}{k}$, che è una somma finita di numeri reali.

(iii)Sia $n \geq 1$ fissato; allora, $Z_{n+1} - Z_n$ è indipendente da $\mathfrak{F}_n$: infatti, $\mathfrak{F}_{Z_{n+1}-Z_n} =$ $=\{\Omega, \{\omega_{n+1=} = D_1\}, \{\omega_{n+1} = D_2\}, \emptyset\}$, che è una σ-algebra indipendente da $\mathfrak{F}_n$, pertanto, fissato un $\Lambda \in \mathfrak{F}_n$, si ha che $E[Z_{n+1} - Z_n|\Lambda] = E[Z_{n+1} - Z_n] = 1 \cdot \psi - \frac{\psi}{1-\psi}(1-\psi) = \psi - \psi = 0$.
Possiamo, di conseguenza, concludere che:

$$P(\Lambda)^{-1}(\int_\Lambda Z_{n+1} dP - \int_\Lambda Z_n dP) = P(\Lambda)^{-1} \int_\Lambda (Z_{n+1} - Z_n) dP$$
$$= E[Z_{n+1} - Z_n|\Lambda] = 0.$$

Esempio 3:
Definizione: *Sia* $(\Omega, \mathfrak{F}, (\mathfrak{F}_\mathfrak{n})_{n \in \mathbb{N}}, P)$ *uno spazio di probabilità filtrato, e sia* $Z = (Z_n)_{n \in \mathbb{N}}$ *un processo stocastico su di esso. Allora* Z *ha incremento indipendente rispetto alla filtrazione* $(\mathfrak{F}_\mathfrak{n})_{n \in \mathbb{N}}$ *se* Z *è adattato a tale filtrazione e se* $\forall n \in \mathbb{N}, n \geq 1$ *e* $\forall 0 \leq k < n$, $Z_n - Z_k$ *è indipendente da* $\mathfrak{F}_k$.

Ora, sia Z un processo stocastico avente incremento indipendente rispetto a $(\mathfrak{F}_\mathfrak{n})_{n \in \mathbb{N}}$, allora:
(i)Se $Z_n \in L^1(\Omega) \forall n \geq 0$, $\check{Z}_n = Z_n - E[Z_n]$ è una martingala.

(ii)Se $Z_n \in L^2(\Omega) \forall n \geq 0$, $Y_n = \check{Z}_n^2 - E[\check{Z}_n^2]$ è una martingala.
(iii)Se $\forall \theta \in \mathbb{R}$, abbiamo che $E[e^{\theta Z_n}] < \infty$,

$$X_n = \frac{e^{\theta Z_n}}{E[e^{\theta Z_n}]}$$

è una martingala.

Ci limiteremo a vedere solamente il primo punto, in quando gli altri due, poi, saranno simili:

$$E[\check{Z}_{n+1}|\mathfrak{F}_n] = E[Z_{n+1} - E[Z_{n+1}]] = E[Z_{n+1} - Z_n + Z_n - E[Z_{n+1}]|\mathfrak{F}_n] =$$

$$= E[Z_{n+1} - Z_n|\mathfrak{F}_n] + E[Z_n - E[Z_{n+1}]|\mathfrak{F}_n] = E[Z_{n+1} - Z_n] + E[Z_n|\mathfrak{F}_n]$$

$$-E[Z_{n+1}|\mathfrak{F}_n] = E[Z_{n+1}] - E[Z_n] + Z_n - E[Z_{n+1}] = Z_n - E[Z_n] = \check{Z}_n.$$

Esempio 4: Sia $\Omega = [a, b]$ un intervallo di $\mathbb{R}$, con $P(A) = |A|(b-a)^{-1}$ per ogni A Boreliano su $[a, b]$ come misura di probabilità, sia $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ una certa filtrazione, e sia dato il processo stocastico $f_n : x \in [a, b] \longmapsto e^{-nx}$. Allora il processo $(f_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ *non* è una martingala: infatti, se lo fosse, si avrebbe che:

$$\int_a^b f_n(x)dx = \int_a^b f_{n-1}(x)dx =, ..., = \int_a^b f_0(x)dx = \int_a^b dx = b - a \forall n \geq 1.$$

Invece:

$$\int_a^b f_n(x)dx = \int_a^b e^{-nx}dx = \frac{e^{-an} - e^{-bn}}{n}.$$

I primi due esempi hanno molti punti in comune: in entrambi, è stata ottenuta una filtrazione da una successione di partizioni sempre più fini di Ω; più specificamente, nel primo esempio, ogni σ-algebra della filtrazione è stata ottenuta dimezzando tutti gli intervalli disgiunti il cui insieme generava la σ-algebra precedente, mentre nel secondo esempio, ogni evento della σ-algebra $\mathfrak{F}_n$ veniva scisso in due eventi tali che il rapporto delle loro probabilità fosse $\frac{\psi}{1-\psi}$, per costruire la σ-algebra $\mathfrak{F}_{n+1}$. In particolare, gli eventi $\Omega^{n,k}$ venivano scomposti in $(\Omega^{n,k} \cap \Omega^{n+1,k}) \cup (\Omega^{n,k} \cap \Omega^{n+1,k+1})$ unione disgiunta di intersezioni di eventi indipendenti; di fatto, è come dividere l'intervallo $[0, 1]$ in due intervalli disgiunti $[0, \psi] \setminus \{\psi\}$ e $[\psi, 1]$. Inoltre, il primo esempio e il secondo, possono essere accomunati osservando che, in generale, se ho uno spazio di probabilità filtrato $(\Omega, \mathfrak{F}, (\mathfrak{F}_\mathfrak{n})_{n\in\mathbb{N}}, P)$ e $Z \in L^1(\Omega)$, allora la costruzione $Z_n = E[Z|\mathfrak{F}_\mathfrak{n}]$ è una martingala (questo è anche il modo più semplice in assoluto di costruire una martingala); nell'esempio 1, $Z(x) = f(x)$, mentre nell'esempio 2, $Z = Z_1$. In un certo senso, l'esempio 2 può essere visto come una versione *discreta* dell'esempio 1, viceversa, l'esempio 1 come una versione *continua* dell'esempio 2.
L'esempio 3 è una dimostrazione di come le martingale possano non essere costruite tutte nello stesso modo.

L'esempio 4 si può generalizzare come segue: data una martingala $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$, dato un $n \in \mathbb{N}$, e dato un $\Lambda \in \mathfrak{F}_n$, a partire da n, la media calcolata su Λ si conserva.

Come si può interpretare il concetto di martingala nella vita reale? Come detto all'inizio del capitolo, essa è la rappresentazione astratta di un cosiddetto *gioco equo*: nell'esempio 2, se h rappresenta il lancio di una monetina ed un giocatore scommette un'unità di denaro sull'uscita del risultato D_1, con la regola che, qualora esca D_2, la perdita sia di $-\frac{\psi}{1-\psi}$ unità di denaro, e Z_n rappresenta la condizione economica in cui il giocatore si trova dopo che sono stati effettuati n lanci, allora, sul lungo termine, la perdita andrà ad eguagliare perfettamente il guadagno (se ad esempio $\psi = \frac{2}{3}$ ed $1-\psi = \frac{1}{3}$, allora $-\frac{\psi}{1-\psi} = -2$, cioè, D_1 ha il doppio di probabilità di uscita rispetto a D_2, ma se esce quest'ultimo, allora la perdita sarà pari al guadagno effettuato con due uscite di D_1). In tutto ciò si può supporre che la monetina non sia lanciata dal giocatore, in quanto egli potrebbe barare, ma da un' entità separata. In ultima analisi, osserviamo che per tutti gli $n \geq 1$, $E[X_n] = 0$: ciò che preme sottolineare, non è tanto il fatto che la media su tutto Ω sia sempre nulla, quanto il fatto che essa **permanga** al variare di n: qui si trova uno dei punti cardine del concetto di martingala.

Esempio 5: L'esempio 2 può essere visto sotto una diversa prospettiva: supponiamo di avere un cosiddetto *processo di Bernoulli*: ossia una successione di v.a. definite su un certo spazio di probabilità Ω, dove $Y_n \sim B_p$, cioè: $Y_n = 1$ con probabilità p e 0 con probabilità $1-p$. Allora, se poniamo $S_n = Y_1 + ... + Y_n$, il processo $S_n - np$ può essere considerato una martingala, se si identifica Ω con lo spazio di probabilità filtrato dell'esempio 2, con D_1 che ha probabilità di uscita p e D_2 che ha probabilità di uscita $1-p$. Qui, diversamente dall'esempio 2, abbiamo che $\forall i \geq 1$, $\omega_i' = 1-p$ se $\omega_i = D_1$, mentre $\omega_i' = -p$ se $\omega_i = D_2$, e $S_n - np$ può essere identificato con $\omega_1' + ... + \omega_n'$, identificando Y_i con ω_i'.

Capitolo 3

Legge dei Grandi Numeri e Teorema del limite centrale.

3.1 Concetti preliminari.

Qui di seguito verranno esposti tutti i concetti e tutti i risultati necessari alla comprensione dell'enunciato e della dimostrazione del *Teorema del Limite Centrale.*

Definizione: *Data una v.a.i.* $X \in L^2$ *definita su un certo spazio di probabilità, chiamiamo varianza di* X *la quantità:*

$$var(X) = E[(X - E[X])^2].$$

Osservazione: Dato uno spazio di probabilità, una v.a. su di esso che stia in L^1, non è detto che si trovi anche in L^2: prendiamo come esempio lo spazio $\Omega = [0,1]$, con $\mathfrak{F} = \mathcal{B}([0,1])$, e $P(A) = M(A)$, $\forall A \in \mathcal{B}([0,1])$ (M indica la Misura di Lebesgue su $\mathbb{R}$). Allora la funzione $f(x) = \dfrac{1}{\sqrt{x}} I_{\Omega \setminus \{0\}} + c I_{\{0\}}(x)$, $c \in \mathbb{R}$, sta in $L^1(\Omega)$; tuttavia, $f(x)^2 = \dfrac{1}{x} I_{\Omega \setminus \{0\}} + c^2 I_{\{0\}}(x)$ non ha integrale finito su $[0,1]$, pertanto $f \notin L^2(\Omega)$. È per questo motivo che nella definizione di varianza, si suppone che la v.a. X stia in L^2 e non in L^1.

Osservazione: $E[(X - E[X])^2] = E[X^2] - E[X]^2$.

Esempio: Sia X una v.a. con distribuzione di Bernoulli, ossia a valori in $\{0,1\}$, e tale che $P(X = 1) = p$ e $P(X = 0) = 1 - p$, con $p \in (0,1)$ (in scrittura: $X \sim B_p$). Allora $E[X] = 1 \cdot p + 0 \cdot (1-p) = p$, e $var(X) = E[X^2] - E[X]^2 = p \cdot 1 + (1-p) \cdot 0 - p^2 = p - p^2 = p(1-p)$.

Definizione: *Data una v.a. X definita su uno spazio di probabilità $(\Omega, \mathfrak{F}, P)$, definiamo la sua funzione cumulativa come:*

$$F_X(x) = P(X \leq x).$$

Definizione: *Diciamo che una successione di v.a. $(X_n)_{n\in\mathbb{N}}$ converge alla v.a. X in distribuzione se:*

$$\exists \lim_{n\longrightarrow+\infty} F_{X_n}(x) = F_X(x)$$

$\forall x \in \mathbb{R}$ tale che $F_X(x)$ sia continua.

Definizione: *Sia X una variabile aleatoria. Allora si chiama densità di probabilità di X quella funzione $p : \mathbb{R} \longrightarrow \mathbb{R}$ non negativa, misurabile ed integrabile (se esiste) tale che:*

$$P(X \in H) = \int_H p(x)dx$$

$\forall H \in \mathcal{B}(\mathbb{R})$.

Osservazione: Se la v.a. è discreta, allora si sostituisce l'integrale con: $\sum_{i\in\mathcal{K}} p(x_i)$, dove $\{x_i : i \in \mathcal{K}\} = X(H)$ ($\mathcal{K}$ è finito o al più numerabile). Nel caso di una v.a. $X \sim B_p$, $p(x) = p^x(1-p)^{1-x}I_{\{0,1\}}(x)$.

Osservazione: $F_X(x) = \int_{-\infty}^{x} p(u)du \ \forall x \in \mathbb{R}$.

Definizione: *Data una v.a. X, diciamo che essa ha distribuzione normale standard se la sua densità di probabilità è data da:*

$$p(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

Scriveremo $X \sim N(0, 1)$.

Osservazione: Se una variabile aleatoria ha distribuzione normale standard, allora essa avrà media (valore atteso) 0 e varianza 1.

Definizione: *Sia X una v.a. con distribuzione $F_X(A)$ e funzione cumulativa $F_X(x)$. Si chiama funzione caratteristica di X la funzione $\Phi_X(t)$ a variabile reale e a valori complessi data da:*

$$\Phi_X(t) = E[e^{itX}] = E[cos(tX)] + iE[sen(tX)].$$

Osservazione: Si può dimostrare che:

$$E[e^{itX}] = \int_{-\infty}^{+\infty} e^{itx} \cdot p(x)dx.$$

Esempio: La funzione caratteristica di una $X \sim B_p$ è uguale a $pe^{it} + 1 - p$. Infatti, se $X \sim B_p$, allora $E[e^{itX}] = \sum_{x\in\mathbb{R}} p(x)e^{itx} = \sum_{x\in\{0,1\}} p(x)e^{itx} = p(0)e^{it\cdot 0} + p(1)e^{it\cdot 1} = (1-p)\cdot 1 + pe^{it} = pe^{it} + 1 - p$.

Osservazione: Una v.a. con distribuzione normale standard avrà sempre funzione caratteristica $e^{-t^2/2}$. Per vederlo, richiamiamo due importanti Teoremi dell'Analisi:

Teorema del cambio di variabili: *Sia $\phi : A \longrightarrow B$ un diffeomorfismo, con A e B aperti di $\mathbb{R}^n$. Sia $f : B \longrightarrow \mathbb{R}$ sommabile. Allora:*

$$\int_B f(y)dy = \int_A f(\phi(x))|detJ_\phi(x)|dx,$$

dove $J_\phi(x) = \left(\dfrac{\vartheta\phi_i(x)}{\vartheta x_j}\right)_{1\le i,j\le n}$ *è la matrice Jacobiana della* ϕ.

Teorema di Fubini: *Sia $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ una funzione sommabile. Esiste allora un insieme $E \subset \mathbb{R}$ di misura nulla (secondo Lebesge) tale che la funzine $x \longrightarrow f(x,y)$ sia sommabile su $\mathbb{R}$ $\forall y \notin E$ e, definita $F(y) = \int_{\mathbb{R}} f(x,y)dx$ se $y \notin E$, $F(y) = 0$ se $y \in E$, $F : \mathbb{R} \longrightarrow \mathbb{R}$, la funzione F sia sommabile e risulti che:*

$$\int_{\mathbb{R}^2} f(x,y)d(x,y) = \int_{\mathbb{R}} F(y)dy.$$

Vista la nullità della misura dell'insieme E, ai fini pratici si può anche scrivere, seppur in maniera non del tutto appropriata:

$$\int_{\mathbb{R}^2} f(x,y)d(x,y) = \int_{\mathbb{R}}[\int_{\mathbb{R}} f(x,y)dx]dy.$$

Adesso calcoliamo: $\int_{-\infty}^{+\infty} e^{-u^2/2}$:
Risulta che $I^2 = \int_{-\infty}^{+\infty} e^{-u^2/2}du \cdot \int_{-infty}^{+\infty} e^{-v^2/2}dv = \int_{-\infty}^{+\infty} e^{-v^2/2}[\int_{-\infty}^{+\infty} e^{-u^2/2}du]dv$ $= \int_{-\infty}^{+\infty}[\int_{-\infty}^{+\infty} e^{-(u^2+v^2)/2}du]dv = (*)$. Adesso applichiamo il Teorema di Fubini: $(*) = \int_{\mathbb{R}^2} e^{-(u^2+v^2)/2}dudv = (**)$. Adesso, avvaliamoci del Teorema del Cambio di Variabili introducendo il diffeomorfismo

$$\phi : (\rho,\theta) \in (0,+\infty) \cup \{0\} \times [0,2\pi] \longmapsto (\rho cos\theta, \rho sen\theta) \in \mathbb{R}^2,$$

in maniera tale da convertire le coordinate (u,v) in coordinate polari. Innanzitutto, $|detJ^\phi(\rho,\theta)| = \rho$, pertanto:

$$(**) = \int_{(0,+\infty)\cup\{0\}\times[0,2\pi]} e^{-\rho^2/2}\rho d\rho d\theta = (***).$$

Richiamiamo quindi in causa il Teorema di Fubini:

$$(***) = \int_0^{2\pi}[\int_0^{+\infty} \rho e^{-\rho^2/2}d\rho]d\theta = \int_0^{2\pi}[-e^{-\rho^2/2}]_0^{+\infty}d\theta = \int_0^{2\pi} d\theta = 2\pi.$$

Dunque, $I^2 = 2\pi$, da cui $I = \sqrt{2\pi}$, poiché la funzione $x \longmapsto e^{-x^2/2}$ è positiva. Ora, se una v.a. X ha distribuzione normale standard, allora:

$$E = E[e^{itX}] = \int_{-\infty}^{+\infty} e^{itx} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx - x^2/2} dx.$$

Adesso, basta notare che $itx - \frac{x^2}{2} = -\frac{(x-it)^2}{2} - \frac{t^2}{2}$, quindi, applicando un normale cambio di variabile, si ottiene il risultato.

La funzione caratteristica ha la proprietà di essere continua e la proprietà di assumere valore 1 in 0. Inoltre, la funzione caratteristica di una somma di v.a. indipendenti è uguale al prodotto delle funzioni caratteristiche di tali variabili.

Teorema d'inversione: *Se X ha funzione caratteristica $\Phi_X(t)$, allora per ogni intervallo (a, b), si ha che:*

$$P(a < X < b) + \frac{P(X = a) + P(X = b)}{2} =$$

$$= \lim_{T \longrightarrow +\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \Phi_X(t) dt.$$

Teorema di unicità: *Se la funzione caratteristica di due variabili aleatorie X, Y è la stessa, allora X ed Y hanno la stessa distribuzione.*

Teorema di continuità: *Siano $(X_n), X$ v.a. tali che: $\lim_{n \longrightarrow +\infty} \Phi_{X_n}(t) = \Phi_X(t)$, $\forall t \in \mathbb{R}$. Allora: $F_{X_n}(x) \longrightarrow F_X(x)$, $\forall x \in \mathbb{R}$ tale che F_X sia continua in x.*

Come ultima cosa, enunciamo il *Teorema di Convergenza Monotona*, che vale in Teoria della Misura, e che quindi si può applicare anche in Teoria della Probabilità, e poi richiamiamo la *Legge Forte dei Grandi Numeri*:

Teorema: *Sia g_n una successione di funzioni definite su uno spazio di misura (Ω, m), m-misurabili, tale che $g_n \nearrow g$ quasi ovunque. Allora:*

$$\int_\Omega g_n dm \nearrow \int_\Omega g dm.$$

In altre parole:

$$\lim_{n \longrightarrow +\infty} \int_\Omega g_n dm = \int_\Omega g dm.$$

Definizione: *Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie. Allora diciamo che esse sono indipendenti ed identicamente distribuite se, presi due indici a piacere i, j distinti, X_i ed X_j sono indipendenti ed hanno la stessa distribuzione.*

Teorema(LGN forte): *Sia* $(X_n)_{n\in\mathbb{N}}$ *una successione di variabili aleatorie indipendenti ed identicamente distribuite, con* $E[X_1^4] < +\infty$ *e* $E[X_n] = \mu$ $\forall n \in \mathbb{N}$. *Allora, posto* $S_n = X_1 + ... + X_n$*, si ha qu.s. che:*

$$\lim_{n\longrightarrow+\infty} \frac{S_n}{n} = \mu.$$

Prima di procedere con la dimostrazione della LGN forte, enunciamo e dimostriamo un Lemma:

Lemma: *Supponiamo che siano verificate le ipotesi della LGN forte; allora esiste una costante* $K < +\infty$ *tale che,* $\forall n \geq 0$,

$$E[(S_n - n\mu)^4] \leq Kn^2.$$

Dimostrazione: Poniamo $Z_i = X_i - \mu$; allora: $(\sum_{i=1}^n Z_i)^4 = \sum_{i=1}^n Z_i^4 + a\sum_{i\neq j} Z_i Z_j^3 + b\sum_{i,j,l distinti} Z_i Z_j Z_l^2 + c\sum_{i,j,l,k distinti} Z_i Z_j Z_l Z_k + d\sum_{i\neq j} Z_i^2 Z_j^2$, con $a, b, c, d \in \mathbb{R}$ opportuni. Ora, per i, j, l, k distinti, abbiamo: $E[Z_i Z_j^3] = E[Z_i]E[Z_j^3] = 0, E[Z_i Z_j Z_l^2] = E[Z_i]E[Z_j]E[Z_l^2] = 0$ e $E[Z_i Z_j Z_l Z_k] = E[Z_i]E[Z_j]E[Z_l]E[Z_k] = 0$; quindi, possiamo prendere in considerazione soltanto l'espressione $\sum_{i=1}^n Z_i^4 + 6\sum_{i\neq j} Z_i^2 Z_j^2$, dove il fatto che $d = 6$ si può dedurre sviluppando l'espressione $(\sum_{i=1}^n Z_i^2 + 2\sum_{i\neq j} Z_i Z_j)^2$. Detto questo, dal momento che $|\{(i,j) : 1 \leq i < j \leq n\}| = \sum_{k=1}^{n-1} k = \frac{n(n-1)}{2}$, risulta:

$$E[\left(\sum_{i=1}^n Z_i\right)^4] = nE[Z_1^4] + 3n(n-1)E[Z_1^2 Z_2^2] = u(n).$$

Osserviamo infine che se $K = 4\max(E[Z_1^4], E[Z_1^2]^2)$, allora $u(n) \leq Kn^2$ $\sharp$.

Dimostrazione della LGN forte: $E[\left(\frac{Sn}{n} - \mu\right)^4] \leq Kn^{-2}$, pertanto:

$$\sum_{n\geq1} E[\left(\frac{Sn}{n} - \mu\right)^4] \leq \sum_{n\geq1} \frac{K}{n^2} = K\sum_{n\geq1} \frac{1}{n^2} < +\infty,$$

quindi: $E[\sum_{k\geq1} \left(\frac{Sn}{n} - \mu\right)^4] < +\infty$, da cui: $\sum_{k\geq1} \left(\frac{Sn}{n} - \mu\right)^4 < +\infty$ qu.s., da cui, segue che $\lim_{n\longrightarrow+\infty} \left(\frac{Sn}{n} - \mu\right)^4 = 0$ qu.s., cioè: $\lim_{n\longrightarrow+\infty} \left(\frac{Sn}{n} - \mu\right) = 0$ qu.s. $\sharp$.

3.2 Enunciato, dimostrazione ed interpretazione.

Teorema del limite centrale: *Sia* $(X_n)_{n\in\mathbb{N}}$ *una successione di variabili aleatorie indipendenti ed identicamente distribuite, con media* 0 *e varianza* 1. *Allora,* *se* $Z \sim N(0,1)$ *ed* $S_n = X_1 + ... + X_n$, $\frac{S_n}{\sqrt{n}} \longrightarrow Z$ *in distribuzione. In altre parole,* $\forall x \in \mathbb{R}$, *si ha che:*

$$\lim_{n \longrightarrow +\infty} P\left(\frac{X_1 + ... + X_n}{\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du.$$

Dimostrazione: Per la dimostrazione, ci avvarremo delle tre disuguaglianze seguenti:

$\diamond$ Se $u \geq 0$, allora $1 \leq e^{-u} - 1 + u \leq \frac{u^2}{2}$.

$\diamond$ Se $t \in \mathbb{R}$, allora $|e^{-it} - 1 - it| \leq \frac{|t|^2}{2}$ e $|e^{-it} - 1 - it + \frac{(it)^2}{2}| \leq \frac{|t|^3}{6}$.

$\diamond$ $\forall x \geq 0$ e $\forall n \in \mathbb{N}$, $e^{-nx} \leq ne^{-x}$.

Sia F la funzione caratteristica di X_n, $\forall n \geq 1$. Allora, $\forall t \in \mathbb{R}$, la funzione caratteristica di $S_n/\sqrt{n}$ è data da:

$$E[e^{itS_n/\sqrt{n}}] = E[e^{it\sum_{k=1}^{n} X_k/\sqrt{n}}] = \left[F\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Di conseguenza, il nostro compito sarà quello di dimostrare che, $\forall t \in \mathbb{R}$,

$$\lim_{n \longrightarrow +\infty} \left[F\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{-t^2/2}.$$

Si noti che, per $t = 0$, non c'è nulla da dimostrare.
Iniziamo la nostra stima, notando che:

$$|\left[F\left(\frac{t}{\sqrt{n}}\right)\right]^n - e^{-t^2/2}| = |\left[F\left(\frac{t}{\sqrt{n}}\right)\right]^n - [e^{-t^2/2n}]^n| \leq$$

$$\leq n|F\left(\frac{t}{\sqrt{n}}\right) - e^{-t^2/2n}|.$$

Il $\leq$ è dato dalla terza disuguaglianza.
Adesso, applichiamo la disuguaglianza triangolare:

$$n|F\left(\frac{t}{\sqrt{n}}\right) - e^{-t^2/2n}| \leq n|F\left(\frac{t}{\sqrt{n}}\right) - (1 - \frac{t^2}{2n})| + n|(1 - \frac{t^2}{2n}) - e^{-t^2/2n}|.$$

Dalla prima disuguaglianza, abbiamo che $n|(1 - \frac{t^2}{2n}) - e^{-t^2/2n}| \leq \frac{n}{2}\left(\frac{t^2}{2n}\right)^2 = \frac{t^4}{8n}$, che tende a 0 al tendere di n all'infinito.

Ora, se X è una variabile aleatoria con funzione caratteristica F, $E[X]=0$ e $E[X^2]=1$, allora $E[itX/\sqrt{n}]=0$ e $E[i^2t^2X^2/2n]=i^2t^2/2n$, pertanto, si può scrivere:

$$n|F\left(\frac{t}{\sqrt{n}}\right)-(1-\frac{t^2}{2n})|=n|E[e^{itX/\sqrt{n}}-(1+\frac{i^2t^2X^2}{2n})]|=$$

$$=n|E[e^{itX/\sqrt{n}}-(1+\frac{itX}{\sqrt{n}}+\frac{i^2t^2X^2}{2n})]|\leq$$

$$\leq nE[|e^{itX/\sqrt{n}}-(1+\frac{itX}{\sqrt{n}}+\frac{i^2t^2X^2}{2n})|]=nE[Y_n],$$

con $Y_n=|e^{itX/\sqrt{n}}-(1+\frac{itX}{\sqrt{n}}+\dfrac{i^2t^2X^2}{2n})|$.

Ora, da una parte, utilizzando la seconda disuguaglianza e la disuguaglianza triangolare, abbiamo che:

$$Y_n\leq|e^{itx/\sqrt{n}}-(1+\frac{itX}{\sqrt{n}})|+\frac{t^2X^2}{2n}\leq\frac{t^2X^2}{2n}+\frac{t^2X^2}{2n}=\frac{t^2X^2}{n}.$$

Dall'altra parte, invece, sempre applicando la seconda disuguaglianza, abbiamo che:

$$Y_n\leq\frac{|t|^3|X|^3}{n^{3/2}}.$$

Sia adesso $\delta>0$ e sia $n\in\mathbb{N}$ fissato.
Laddove $|X|\leq\delta\sqrt{n}$, sfruttiamo il fatto che $Y_n\leq\frac{|t|^3|X|^3}{6n^{3/2}}$, dicendo che:

$$Y_n\leq\frac{|t|^3|X|^3}{6n^{3/2}}\leq\frac{|t|^3X^2\delta\sqrt{n}}{6n^{3/2}}=\frac{|t|^3X^2\delta}{6n}.$$

Da qui, possiamo dire che:

$$E[\frac{|t|^3|X|^3I_{|X|\leq\delta\sqrt{n}}}{6n^{3/2}}]\leq E[\frac{|t|^3X^2I_{|X|\leq\delta\sqrt{n}}\delta}{6n}]\leq\frac{|t|^3\delta}{6n}.$$

Sia dato $\epsilon>0$: allora nulla ci vieta di scegliere δ tale da soddisfare $\frac{|t|^3\delta}{6}\leq\frac{\epsilon}{2}$.
Ora, per tale $\delta>0$, consideriamo la successione di v.a. $t^2X^2I_{|X|\leq\delta\sqrt{n}}$: allora, tale successione converge in maniera crescente a t^2X^2, pertanto, applicando il Teorema di Convergenza Monotona, si ottiene che:

$$E[t^2X^2I_{|X|\leq\delta\sqrt{n}}]\nearrow E[t^2X^2]=t^2.$$

Dunque, si può scegliere un $N\in\mathbb{N}$ tale che, $\forall n\geq N$:

$$t^2E[X^2I_{|X|>\delta\sqrt{n}}]=t^2(1-E[X^2I_{|X|\leq\delta\sqrt{n}}])\leq\frac{\epsilon}{2}.$$

Quindi, si può concludere dicendo che $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$:

$$nE[Y_n] \leq t^2 E[X^2 I_{|X|>\delta\sqrt{n}}] + nE[\frac{|t|^3|X|^3 I_{|X|<=\delta\sqrt{n}}}{6n^{3/2}}] \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \forall n \geq N,$$

da cui:

$$|\left[F\left(\frac{t}{\sqrt{n}}\right)\right]^n - e^{-t^2/2}| \leq \epsilon$$

$\forall n \geq N$ $\sharp$.

Che interpretazione si può dare al TLC? Analizziamo la situazione in cui una monetina viene lanciata ripetute volte, e i due risultati complementari hanno uno probabilità $p \in (0,1)$ e l'altro probabilità $1-p$, e costruiamo il processo stocastico X_n, dove X_n(n-esimo lancio)=1 con probabilità p e 0 con probabilità $1-p$. Allora, poiché i lanci sono tutti indipendenti, lo saranno anche le X_n; inoltre, $Xn \sim B_p$, $\forall n \geq 1$. Ora, $\forall n \in \mathbb{N}$, avremo che $E[X_n] = p$ e $var(X_n) = p(1-p)$; ebbene, se $S_n = X_1 + ... + X_n$, la Legge Forte dei Grandi Numeri ci dice che $\lim_{n \longrightarrow +\infty} \frac{S_n - np}{n} = 0$ qu.s., cioè, $\frac{S_n}{n} \longrightarrow p$ al tendere di n a $+\infty$. In parole povere, alla lunga, il rapporto fra il numero di volte in cui è uscito il risultato con probabilità p ed il numero di lanci effettuati fino a quel momento è destinato a stabilizzarsi progressivamente a p. Un'altra interpretazione equivalente che si può dare alla LGN forte, è che la differenza fra la media del guadagno effettivo che i lanci effettuati fino ad un certo momento hanno comportato e la media teorica sarà destinata, sul lungo termine, ad azzerarsi. In termini probabilistici, abbiamo che $P\left(\lim_{n \longrightarrow +\infty} \frac{S_n - pn}{n} = 0\right) = 1$. Anche il TLC ci dà informazioni su come si comporta un fenomeno dopo che è stato fatto un grande numero di tentativi: nel caso del lancio della monetina, la variabile aleatoria $\frac{S_n - np}{\sqrt{p(1-p)}}$ ha media 0 e varianza 1, quindi:

$$\lim_{n \longrightarrow +\infty} P\left(\frac{S_n - pn}{\sqrt{pn(1-p)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du,$$

$\forall x \in \mathbb{R}$.

Innanzitutto, osserviamo che $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$ non dipende da p; questo significa che i fenomeni aleatori simili al lancio di una monetina tendono, sul lungo termine, a comportarsi nella stessa maniera. Si può poi osservare che $\frac{e^{-u^2/2}}{\sqrt{2\pi}}$ è una funzione simmetrica: l'evento in cui il guadagno effettivo dopo n-lanci si discosta in eccesso da quello teorico ha la stessa probabilità dell'evento in cui il guadagno effettivo si discosta in difetto, ossia 1/2. Andando più in profondità, si può dire che il TLC sia una rappresentazione in formule del modo in cui la difficoltà di discostarsi dal guadagno teorico, in eccesso o in difetto, varia all'aumentare di tale discostamento: si tratta di una difficoltà sempre più alta, che però cresce sempre più debolmente, e la funzione $x \longmapsto \int_{-\infty}^{x} \frac{e^{-u^2/2}}{\sqrt{2\pi}}$ ci fornisce informazioni su come questa crescita avvenga. In ultima analisi, possiamo osservare che il risultato del TLC, applicato all'esempio che stiamo prendendo in considerazione, deriva dal fatto che la funzione caratteristica di $\frac{X_n - p}{\sqrt{p(1-p)}}$ tende alla funzione caratteristica di una $Z \sim N(0,1)$, quindi la funzione cumulativa di $\frac{X_n - p}{\sqrt{p(1-p)}}$ tende alla stessa funzione cumulativa di Z: di conseguenza, sappiamo anche come stimare la probabilità che il guadagno effettuato si trovi entro una certa gamma di valori. Il TLC, ci fornisce di conseguenza una quantità di informazioni maggiore rispetto alla LGN forte, la quale ci dice semplicemente che sul lungo termine, una sequenza di ripetizioni dello stesso fenomeno aleatorio si assesterà su un comportamento dato.

3.3 Il TLC dal punto di vista delle martingale.

Definizione: *Diciamo che una successione di variabili aleatorie X_n, con $n \geq 1$, converge in probabilità ad una v.a. X se $\lim_{n \longrightarrow +\infty} P(|X_n - X| \geq \epsilon) = 0 \ \forall \epsilon > 0$.*

Definizione: *Una successione X_n, con $n \geq 1$, è detta uniformemente limitata se $\exists k > 0 : P(|X_n| \geq k) = 0 \ \forall n \geq 1$.*

Definizione: *Una famiglia di variabili aleatorie del tipo $X_{n,m}$ con $n \geq 1$ e $1 \leq m \leq n$ si chiama vettore triangolare.*

Il fatto di chiamare la famiglia $X_{n,m}$ *vettore triangolare* deriva dal fatto che essa può essere rappresentata come segue:

$X_{1,1}$
$X_{2,1}$ $X_{2,2}$
$X_{3,1}$ $X_{3,2}$ $X_{3,3}$
.
.
.
.

Definizione: *Sia dato uno spazio di probabilità filtrato $(\Omega, \mathfrak{F}, (\mathfrak{F}_n)_{n\in\mathbb{N}}, P)$. Allora, un processo stocastico $X = (X_n)_{n\geq 1}$ è detto differenza di martingala rispetto alla filtrazione $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ se :*
(1) X è adattato a $(\mathfrak{F}_n)_{n\in\mathbb{N}}$.
(2) $\forall n \geq 1$, X_n è integrabile e $E[|X_n|] < +\infty$.
(3) $E[X_{n+1}|\mathfrak{F}_n] = 0 \ \forall n \in \mathbb{N}$.

Consideriamo ora un vettore triangolare $X = \{X_{n,m}, \mathfrak{F}_{n,m}, \mathfrak{F}_{n,0}\}_{n\geq 1, 1\leq m\leq n}$ che goda delle proprietà seguenti: $(\mathfrak{F}_{n,i})_{0\leq i\leq n}$ è una filtrazione, le $X_{n,1},...,X_{n,n}$ stanno in L^2, e $X_{n,i}$ è $\mathfrak{F}_{n,i}$-misurabile, e $E[X_{n,i}|\mathfrak{F}_{n,i-1}]=0$.Per un motivo che vedremo più avanti, poniamo $S_n^*=X_{n,1}+...+X_{n,n}$. Allora chiamiamo X vettore triangolare di differenze di martingala. Detto ciò, definiamo:

$$\sigma_{n,m}^2=E[X_{n,m}^2|\mathfrak{F}_{n,m-1}],$$

$$V_{n,m}=\sigma_{n,1}^2+...+\sigma_{n,m}^2,$$

$$V_n=V_{n,n}.$$

Osserviamo che $\sigma_{n,m}^2$ e $V_{n,m}$ sono $\mathfrak{F}_{n,m-1}$-misurabili.
Ora, sia $L_n(\epsilon)=\sum_{m=1}^n E[X_{n,m}^2 I_{|X_{n,m}|\geq\epsilon}|\mathfrak{F}_{n,m-1}]$ $\forall\epsilon>0$, allora il suo valore atteso è dato da: $E[L_n(\epsilon)]=\sum_{m=1}^n E[[X_{n,m}^2 I_{|X_{n,m}|\geq\epsilon}]$. Diciamo che X gode della Condizione di Lindeberg-Feller se $\forall\epsilon>0$, $\lim_{n\longrightarrow+\infty}L_n(\epsilon)=0$ in probabilità. Si può dimostrare che la condizine $\lim_{n\longrightarrow+\infty}E[L_n(\epsilon)]=0$ è più forte, e che le due condizioni diventano equivalenti quando la successione V_n è uniformemente limitata.

Lemma 1: *Se la Condizione di Lindeberg-Feller è soddisfatta, allora:*

$$\lim_{n\longrightarrow+\infty}\max_{k\leq n}\sigma_{n,k}^2=0$$

in probabilità.

Diciamo che X gode della Condizione di Stabilità se $\lim_{n\longrightarrow+\infty}V_n=\eta$ in probabilità per un certo η reale.

Lemma 2: *Supponiamo che la Condizione di Stabilità sia verificata, e poniamo $X'_{n,m}=X_{n,m}I_{V_{n,m}\leq\eta+1}$. Allora $X'=\{X_{n,m},\mathfrak{F}_{n,m},\mathfrak{F}_{n,0}\}_{n\geq1\wedge1\leq m\leq n}$ è un vettore triangolare di differenze di martingala, dove $\lim_{n\longrightarrow+\infty}V'_n=\eta+1$ in probabilità. Di più, se X soddisfa la Condizine di Lindeberg-Feller, allora anche X' la soddisfa.*

Adesso procediamo con l'enunciato e la dimostrazione del *Teorema del Limite Centrale per Martingale:*

Teorema(TLCM): *Se su X sono soddisfatte sia la Condizione di Lindeberg-Feller, sia la Condizione di Stabilità, con $\eta=1$, allora:*

$$S_n^*\longrightarrow Z\sim N(0,1)$$

in distribuzione.

Dimostrazione: Lo dimostreremo sotto l'ipotesi che $V_n\leq2$ quasi sicuramente $\forall n\geq1$. Allora, il nostro compito sarà quello di dimostrare che $\lim_{n\longrightarrow n}E[e^{itS_n^*}]=e^{-t^2/2}$ $\forall t$. Scriviamo:

$$E[e^{itS_n^*}]-e^{-t^2/2}=e^{-t^2/2}E[R_{n,1}(t)+R_{n,2}(t)],$$

dove $R_{n,1}(t) = e^{itS_n^*}[e^{t^2/2} - e^{V_n t^2/2}]$ e $R_{n,2}(t) = e^{tS_n^*} + e^{V_n t^2/2} + 1$.
Chiaramente, $E[|R_{n,1}|] \leq E[e^{t^2/2} - e^{V_n t^2/2}] \longrightarrow 0$, poiché $V_n \longrightarrow 1$ in probabilità e le V_n sono limitate, e quindi integrabili. Per stimare il secondo termine, scriviamo, ponendo $S_{n,k} = X_{n,1} + ... + X_{n,k}$:

$$R_{n,2}(t) = \sum_{k=1}^{n}[e^{itS_{n,k}+V_{n,k}t^2/2} - e^{itS_{n,k-1}+V_{n,k-1}t^2/2}] =$$

$$= \sum_{k=1}^{n} e^{itS_{n,k-1}+V_{n,k}t^2/2}[e^{itX_{n,k}} - e^{-\sigma_{n,k}^2 t^2/2}].$$

Qui,

$$|E[e^{itS_{n,k-1}+V_{n,k}t^2/2}[e^{itX_{n,k}} - e^{-\sigma_{n,k}^2 t^2/2}]| = |E[e^{itS_{n,k-1}+V_{n,k}t^2/2} E[e^{itX_{n,k}} - e^{-\sigma_{n,k}^2 t^2/2}|\mathfrak{F}_{n,k-1}]|$$
$$\leq e^{t^2} E[E[e^{itX_{n,k}} - e^{-\sigma_{n,k}^2 t^2}|\mathfrak{F}_{n,k-1}]].$$

Adesso:

$$e^{itX_{n,k}} = 1 + itX_{n,k} - \frac{t^2 X_{n,k}^2}{2} + R_3(tX_{n,k}),$$

mentre:

$$e^{-\sigma_{n,k}^2 t^2/2} = 1 - \frac{\sigma_{n,k}^2 t^2}{2} + r_2\left(\frac{\sigma_{n,k}^2 t^2}{2}\right),$$

dove $|R_3| \leq min(t^2 X_{n,k}^2, \frac{|t^3 X_{n,k}^3|}{6})$ e $|R_2| \leq \frac{\sigma_{n,k}^4 t^4}{4}$.

Ora, osservando che: $E[e^{itX_{n,k}} - e^{-\sigma_{n,k}^2/2}|\mathfrak{F}_{n,k-1}] = E[R_3(tX_{n,k}) - r_2\left(\frac{\sigma_{n,k}^2 t^2}{2}\right)|\mathfrak{F}_{n,k-1}]$,
alla luce delle disuguaglianze appena viste, risulta:

$$|E[R_{n,2}(t)]| \leq e^{t^2}\sum_{k=1}^{n} E[|R_3(tX_{n,k})| + |r_2(\sigma_{n,k}^2)/2|].$$

Qui, $\sum_{k=1}^{n} E[|r_2(\sigma_{n,k}^2)/2|] \leq \frac{1}{4}t^4 E[max_{k\leq n}\sigma_{n,k}^2 V_n] \longrightarrow 0$ in probabilità per $n \longrightarrow +\infty$, giacché $\max_{k\leq n}\sigma_{n,k}^2 V_n \longrightarrow 0\cdot\eta = 0$ in probabilità e $\max_{k\leq n}\sigma_{n,k}^2 V_n^{2\leq 4}$ quasi ovunque, $\forall n \geq 1$.
Ora, sia $\epsilon > 0$; allora, possiamo dire che:

$$E[|R_3(tX_{n,k})|] \leq \frac{|t^3|}{6}E[|X_{n,k}|^3 I_{|X_{n,k}|\leq\epsilon}] + t^2 E[|X_{n,k}|^2 I_{|X_{n,k}|>\epsilon}]$$
$$\leq \frac{\epsilon t^3}{6}E[|X_{n,k}|^2 I_{|X_{n,k}|\leq\epsilon}] + t^2 E[|X_{n,k}|^2 I_{|X_{n,k}|>\epsilon}].$$

Da qui, si ottiene che $\sum_{k=1}^{n} E[|R_3(tX_{n,k})|] \leq\leq \frac{1}{6}t^3 E[V_n] + t^2 E[L_n(\epsilon)]$, ragion per cui:

$$\limsup_{n\longrightarrow+\infty}\sum_{k=1}^{n} E[|R_3(tX_{n,k})|] \leq \frac{\epsilon t^2}{3}.$$

Quindi, segue che $\lim_{n \longrightarrow +\infty} |E[R_{n,2}|] = 0\sharp$.

Corollario: *Sia data una differenza di martingala $X_1, X_2, ...$ ad elementi in L^2 con filtrazione $\mathfrak{F}_0 \leq \mathfrak{F}_2 \leq \mathfrak{F}_2...$ e sia $S'_n = X_1 + ... + X_n$. Allora, se:*

$$\frac{1}{n}\sum_{k=1}^{n} E[X_k^2|\mathfrak{F}_{k-1}] \longrightarrow 1$$

in probabilità, e

$$\frac{1}{n}\sum_{k=1}^{n} E[X_k^2 I_{|X| \geq \epsilon\sqrt{n}}|\mathfrak{F}_{k-1}] \longrightarrow 0$$

in probabilità $\forall\epsilon > 0$, $\dfrac{S'_n}{\sqrt{n}} \longrightarrow Z \sim N(0,1)$ *in distribuzione.*

Dimostrazione: Basta applicare il Teorema per $X_{n,k} = \dfrac{X_k}{\sqrt{n}}$ e $\mathfrak{F}_{n,i} = \mathfrak{F}_i$ $\sharp$.

Adesso, da una martingala $(Z_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ si può sempre ricavare una differenza di martingala: basta porre $X_k = Z_k - Z_{k-1}$ per $k \geq 1$. Ora, supponiamo di avere un processo stocastico $X_k \sim B_p$; allora il processo $S'_n = S_n - np$ è una martingala, se identifichiamo gli X_k con gli Y_k $(k = n)$ dell'esempio 5 visto a pagina 21. Ora, se aggiungiamo un $X_0 \sim I_{\emptyset}$, il processo $W_{k+1} = \dfrac{S_{k+1} - S_k}{\sqrt{p(1-p)}} = \dfrac{X_{k+1} - p}{\sqrt{p(1-p)}}$ è una differenza di martingala, dove:

(1) $E[W_{k+1}^2|\mathfrak{F}_k] = E[W_{k+1}^2] = 1$ $\forall k \geq 0$.
(2) $\forall\epsilon > 0$, si ha che: $E[W_k^2 \cdot I_{|W_k| \geq \epsilon\sqrt{n}}|\mathfrak{F}_{k-1}] = E[W_k^2 \cdot I_{|W_k| \geq \epsilon\sqrt{n}}] = 0$ per ogni $n > \left(\dfrac{1-p}{\epsilon}\right)^2$.
Quindi, valgono le ipotesi del Corollario appena visto.

Capitolo 4

Grandi Deviazioni e Legge del Logaritmo Iterato(LLI).

Tutti i ragionamenti che seguono nelle due sezioni di questo capitolo, sono fatti sotto l'ipotesi di avere un processo di Bernoulli $X_n \sim B_p$, con $n \in \mathbb{N}$ e $p \in (0,1)$.

4.1 Le grandi deviazioni.

Teorema delle grandi deviazioni(TGD): *Sia data una successione di v.a.* $(X_n)_n$, *con* $X_n \sim B_p$, $p \in (0,1)$, $\forall n \geq 1$. *Allora, posto* $S_n = \sum_{k=1}^{n} X_n$, $\forall 0 < \epsilon < 1-p$ *e* $\forall n \geq 1$, *si ha che:*

$$P\left(\frac{S_n}{n} \geq p+\epsilon\right) \leq e^{-nh_+(\epsilon)},$$

dove $h_+(\epsilon) = (p+\epsilon)\log\left(\frac{p+\epsilon}{p}\right) + (1-p-\epsilon)\log\left(\frac{1-p-\epsilon}{1-p}\right)$.

In pratica, si tratta di una disuguaglianza volta a stimare la probabilità che dopo un certo numero di lanci, la media del guadagno effettivo si discosti in eccesso almeno di un $\epsilon \in (0, 1-p)$ dalla media teorica del guadagno.

Osservazione: $h_+(\epsilon)$ è una funzione crescente su $(0, 1-p)$. Infatti, $h'_+(\epsilon) = \log\left(\frac{(p+\epsilon)(1-p)}{p(1-p-\epsilon)}\right)$, e maggiorando tale derivata a 0, si ottiene: $\log\left(\frac{(p+\epsilon)(1-p)}{p(1-p-\epsilon)}\right) \geq 0 = \log 1$, da cui $\frac{(p+\epsilon)(1-p)}{p(1-p-\epsilon)} \geq 1$, da cui $p - p^2 + \epsilon - \epsilon p \geq p - p^2 - \epsilon p$, cioè $\epsilon \geq 0$. Inoltre, $\forall n \in \mathbb{N}$, $(e^{-nh_+(\epsilon)})' = -nh'_+(\epsilon)e^{-nh'_+(\epsilon)} \leq 0$ $\forall \epsilon \in (0, 1-p)$, pertanto $e^{-nh_+(\epsilon)}$ è decrescente su $(0, 1-p)$ $\forall n \in \mathbb{N}$.

Prima di procedere con la dimostrazione del TGD, ci servirà conoscere un importante enunciato:

Disuguaglianza di Markov: *Per una v.a.* X *non negativa, dove* $X(s) \geq 0$ $\forall s \in \Omega$*, allora,* $\forall a > 0$*, si ha che:*

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Dimostrazione: $A = \{s \in \Omega | X(s) \geq a\}$. Allora:

$$E[X] = \sum_{s \in \Omega} P(X = s)X(s) = \sum_{s \in A} P(X = s)X(s) + \sum_{s \notin A} P(X = s)X(s) \geq$$

$$\geq \sum_{s \in A} P(X = s)X(s) \geq \sum_{s \in A} a \cdot P(X = s) = a \cdot \sum_{s \in A} P(X = a) = aP(A).$$

Da qui: $\frac{E[X]}{a} \geq a \cdot P(A)$, in altre parole, $\frac{E[X]}{a} \geq P(A)$, che è esattamente ciò che volevamo far vedere.♯

Dimostrazione del TGD: Cominciamo con l'osservare che se $\frac{S_n}{n} \geq p + \epsilon$, allora $S_n - np - n\epsilon \geq 0$, da cui $t(S_n - np - n\epsilon) \geq 0$ $\forall t > 0$, quindi $e^{t(S_n - np - n\epsilon)} \geq 1$ $\forall t > 0$. Viceversa, se $e^{t(S_n - np - n\epsilon)} \geq 1$ $\forall t > 0$, allora $t(S_n - np - n\epsilon) \geq 0$ $\forall t > 0$, pertanto $S_n \geq np + n\epsilon$, dunque $\frac{S_n}{n} \geq p + \epsilon$. Si può, di conseguenza, concludere che:

$$P\left(\frac{S_n}{n} \geq p + \epsilon\right) = P(e^{t(S_n - np - n\epsilon)} \geq 1 \forall t > 0).$$

Il "$\forall t > 0$" si può, in realtà, tirare fuori.

Adesso, applichiamo la Disuguaglianza di Markov sulla variabile aleatoria $e^{t(S_n - np - n\epsilon)}$:

$$P(e^{t(S_n - np - n\epsilon)} \geq 1) \leq E[e^{t(S_n - np - n\epsilon)}] = e^{-nt(p+\epsilon)} E[e^{tS_n}] =$$

$$= e^{-nt(p+\epsilon)} \sum_{k=0}^{n} e^{tk} P(S_n = pk) = e^{-nt(p+\epsilon)} \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} =$$

$$= e^{-nt(p+\epsilon)} \sum_{k=0}^{n} (pe^t)^k (1-p)^{n-k} \binom{n}{k} = e^{-nt(p+\epsilon)} (1 - p + pe^t)^n =$$

$$= e^{-n(t(p+\epsilon) - \log(1-p+pe^t))}.$$

Poniamo adesso $g(t) = t(p + \epsilon) - \log(1 - p + pe^t)$. Allora $g(0) = 0$ e $g'(t) = p + \epsilon - pe^t(1 - p + pe^t)^{-1}$, da cui $g'(0) = \epsilon$. Inoltre, $\lim_{t \longrightarrow +\infty} g'(t) = p + \epsilon - 1 < 0$. Da qui, il *sup* si ottiene per un qualche valore di t strettamente positivo. Detto questo, la derivata di $g(t)$ è nulla solo in

$$s = \log\left(\frac{-p + p^2 - \epsilon + \epsilon p}{p(p + \epsilon - 1)}\right) = \log\left(\frac{(p + \epsilon)(1 - p)}{p(1 - p - \epsilon)}\right).$$

Infatti, risolvendo l'equazione $p + \epsilon - pe^t(1-p+pe^t)^{-1} = 0$, viene: $p + \epsilon - \dfrac{pe^t}{1-p+pe^t} = 0$, quindi $\dfrac{(p+\epsilon)(1-p+pe^t) - pe^t}{1-p+pe^t} = 0$, da cui $p - p^2 + p^2e^t + \epsilon - \epsilon pe^t - pe^t = 0$, dunque $(p^2 + \epsilon p - p)e^t = p^2 - p - \epsilon + \epsilon p = (1-p)(p+\epsilon)$, pertanto $e^t = \dfrac{(1-p)(p+\epsilon)}{p(1-p-\epsilon)}$. Perciò, $g(t)$ ha valore massimo $h_+(\epsilon)$, poiché:

$$g(s) = (p+\epsilon)\log\left(\frac{p+\epsilon}{p}\right) + (p+\epsilon)\log\left(\frac{1-p}{1-p-\epsilon}\right)$$

$$-\log\left(1-p+p\left(\frac{(p+\epsilon)(1-p)}{p(1-p-\epsilon)}\right)\right) = (p+\epsilon)\log\left(\frac{p+\epsilon}{p}\right) +$$

$$(p+\epsilon)\log\left(\frac{1-p}{1-p-\epsilon}\right) - \log\left((1-p)\left(1+\frac{p+\epsilon}{1-p-\epsilon}\right)\right) =$$

$$(p+\epsilon)\log\left(\frac{p+\epsilon}{p}\right) + (p+\epsilon)\log\left(\frac{1-p}{1-p-\epsilon}\right) - \log\left(\frac{1-p}{1-p-\epsilon}\right) =$$

$$= h_+(\epsilon)\sharp.$$

Corollario 1: $\forall\epsilon \in (0,p)$ e $\forall n \geq 1$, abbiamo:

$$P\left(\frac{S_n}{n} \leq p - \epsilon\right) \leq e^{-nh_-(\epsilon)},$$

dove $h_-(\epsilon) = h_+(-\epsilon)$.

Dimostrazione: Indichiamo con S_n^c il processo complementare ad S_n, con X_n^c(n-esimo lancio)$= 1$ con probabilità $1-p$ e 0 con probabilità p, ed indichiamo con $h_+^c(\epsilon)$ l'analogo rispetto ad S_n^c della $h_+(\epsilon)$. Pertanto: $P\left(\dfrac{S_n^c}{n} \geq 1-p+\epsilon\right) \leq e^{-nh_+^c(\epsilon)}$.
Osserviamo infine che $h_+^c(\epsilon) = (1-p-\epsilon)\log\left(\dfrac{1-p-\epsilon}{1-p}\right) + (p-\epsilon)\log\left(\dfrac{p-\epsilon}{p}\right) = h_+(-\epsilon) = h_-(\epsilon)$.
La disuguaglianza diventa dunque:

$$P\left(1 - \frac{S_n^c}{n} \leq p - \epsilon\right) = P\left(\frac{S_n}{n} \leq p - \epsilon\right) \leq e^{-nh_-(\epsilon)}\sharp.$$

Corollario 2: $\forall\epsilon \in (0, min(p, 1-p))$ e $\forall n \geq 1$, si ha che:

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq e^{-nh_+(\epsilon)} + e^{-nh_-(\epsilon)}.$$

Come ultima cosa, diamo l'enunciato di un importante risultato che prende il nome di *Teorema delle Piccole Deviazioni*:

Teorema: *Supponiamo che:*
(1) a_n *sia una successione di numeri reali.*
(2) $a_n \longrightarrow +\infty$ *per* $n \longrightarrow +\infty$.
(3) $\lim_{n \longrightarrow +\infty} \frac{a_n}{n^{1/6}} = 0$.
Allora:

$$P\left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}\frac{a_n}{n^{1/6}}\right) \sim \frac{1}{a_n\sqrt{2\pi}}e^{-a_n^2/2}.$$

Osserviamo che, utilizzando de l'Hôpital, possiamo calcolare il seguente limite:

$$\lim_{x \longrightarrow +\infty} \frac{(\sqrt{2\pi})^{-1}\int_x^{+\infty} e^{-t^2/2}dt}{(x\sqrt{2\pi})^{-1}e^{-x^2/2}} = \lim_{x \longrightarrow +\infty} \frac{-e^{-x^2/2}}{-\frac{x}{x}e^{-x^2/2} + \frac{1}{x^2}e^{-x^2/2}} = \lim_{x \longrightarrow +\infty} \frac{-1}{-1 + \frac{1}{x^2}} =$$

$$= 1.$$

Questo significa che, per $n \longrightarrow +\infty$, $\frac{1}{a_n\sqrt{2\pi}}e^{-a_n^2/2} \sim \frac{1}{\sqrt{2\pi}}\int_{a_n}^{+\infty} e^{-t^2/2}dt$, cioè, per $n \longrightarrow +\infty$, la successione $\frac{1}{a_n\sqrt{2\pi}}e^{-a_n^2/2}$ si avvicina ad essere l'area della coda di una curva Gaussiana, poiché il grafico della funzione $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ è una curva Gaussiana.

4.2 Enunciato, dimostrazione ed interpretazione.

Chiamiamo limite superiore di una successione a_n, il numero $inf\,\{M|\exists n_0 : a_n \leq M \forall n \geq n_0\}$ e limite inferiore il numero $sup\,\{m|\exists n_0 : a_n \geq m \forall n \geq n_0\}$.

Teorema(LLI di Khinchin): *Quasi sicuramente,*

$$\limsup_{n \longrightarrow +\infty} \frac{S_n - np}{\sqrt{2np(1-p)\log\log(n)}} = 1,$$

$$\liminf_{n \longrightarrow +\infty} \frac{S_n - np}{\sqrt{2np(1-p)\log\log(n)}} = -1.$$

Prima di precedere con la dimostrazione, enunciamo e dimostriamo una serie di risutati preliminari e non; poniamo, per alleggerire le notazioni, $\alpha(n) = \sqrt{2p(1-p)n\log\log n}$.

Stima di Haudsorff: $\forall \epsilon > 0$,

$$S_n - np = o(n^{\epsilon+1/2})$$

per $n \longrightarrow +\infty$.

Cioè, $\lim_{n \longrightarrow +\infty} \frac{S_n - np}{n^{\epsilon+1/2}} = 0$ qu.s. per ogni $\epsilon > 0$.

Dimostrazione: $\forall n \in \mathbb{N}$, poniamo $R_n = S_n - np = \sum_{i=1}^{n} X'_i$, dove $X'_i = X_i - p$. Sia $k \in \mathbb{N}$, $k > 0$ fissato e cerchiamo di dare una stima di $E[R_n^{2k}]$. Innanzitutto, R_n^{2k} è una somma di prodotti del tipo $X'_{i_1} \cdots X'_{i_{2k}}$, dove $\{i_v\}_{1 \leq v \leq 2k}$ è un insieme di indici apparenenti a $\{1, ..., n\}$, che non sono necessariamente a due a due distinti. Cioè, ogni prodotto $X'_{i_1} \cdots X'_{i_{2k}}$ scaturisce da una funzione che ha $\{1, ..., 2k\}$ come dominio e $\{1, ..., n\}$ come codominio. Adesso, se $j \in \{1, ..., n\}$ appare solo una volta nel prodotto $X'_{i_1} \cdots X'_{i_{2k}}$, allora $E[X'_{i_1} \cdots X'_{i_{2k}}] = 0$ per via dell'indipendenza delle variabili aleatorie. Si noti, inoltre, che per tutti gli insiemi di indici, $E[X'_{i_1} \cdots X'_{i_{2k}}] \leq 1$. Questo si può dimostrare per induzione: innanzitutto, se $X \sim B_p$, allora $E[X - p] = 0$ e $E[(X - p)^2] = p(1 - p) \leq 1$. Ora, poiché $p \in (0, 1)$, allora $\forall n \in \mathbb{N}$, $1 + p^{n-1} - p^n = 1 + p^{n-1}(1 - p) > 0$, da cui $-p(1 + p^{n-1} - p^n) \leq 0$, pertanto $(1 - p)(1 + p^n) = 1 - p - p^n + p^{n+1} \leq 1$. Considerando che $(1-p)(1-p^n) \leq 1$, allora si può dire che $(1-p)(1-(-p)^n) \leq 1$ $\forall n \in \mathbb{N}$. Alla luce di quanto appena visto, partiamo ora dal presupposto che, per un certo $n \in \mathbb{N}$, $E[(X - p)^n] \leq 1$; allora, utilizzando il fatto che $\forall n \in \mathbb{N}$, $X^n = 1$ con probabilità p e 0 con probabilità $1 - p$, e quindi la media non cambia, risulta che:

$$E[(X - p)^{n+1}] = E[(X - p)(X - p)^n] = E[X(X - p)^n] - pE[(X - p)^n] =$$

$$= E[(X - p)^n] - (-p)^n + (-p)^n p - pE[(X - p)^n] = (1 - p)(E[(X - p)^n] - (-p)^n)$$

$$\leq (1 - p)(1 - (-p)^n) \leq 1.$$

Da qui:

$$E[R_n^{2k}] = \sum_{1 \leq i_1, ..., i_{2k} \leq n} E[X'_{i_1} \cdots X'_{i_{2k}}] \leq N(k, n),$$

dove $N(k, n)$ è il numero di funzioni da $\{1, ..., 2k\}$ a $\{1, ..., n\}$, che prendono ogni valore almeno due volte. Sia $M(k)$ il numero di partizioni di $\{1, ..., 2k\}$ in sottoinsiemi che contengano almeno due elementi. Allora, se P è una di tali partizioni, P contiene al più k elementi. Il numero di funzioni che sono costanti su ogni insieme di P è al più n^k. Pertanto, $N(k, n) \leq n^k M(k)$. Ora, sia $\epsilon > 0$, e consideriamo: $E[(n^{-\epsilon-1/2} R_n)^{2k}] \leq n^{-2k\epsilon-k} N(k, n) \leq n^{-2k\epsilon} M(k)$.

Sia $k > \frac{1}{2\epsilon}$. Allora: $\sum_{n \geq 1} E[(n^{-\epsilon-1/2} R_n)^{2k}] < +\infty$. Ora, in generale, se $\sum_{n=1}^{+\infty} E[|Y_n|]$ converge, allora la successione di variabili aleatorie Y_n tende alla variabile aleatoria identicamente nulla quasi sicuramente. Pertanto, $(n^{-\epsilon-1/2} R_n)$ $\longrightarrow 0$ quasi sicuramente per $n \longrightarrow +\infty$. Questo significa che $\forall \epsilon > 0$, esiste un evento trascurabile (ossia di probabilità nulla), dipendente da ϵ, all'infuori del quale $n^{-\epsilon-1/2} R_n \longrightarrow 0$. Detto ciò, consideriamo un insieme numerabile di valori di ϵ che tendono a 0. Poiché un'unione numerabile di eventi trascurabili è ancora un evento trascurabile, risulta che $\forall \epsilon > 0$, $n^{-\epsilon-1/2} R_n \longrightarrow 0$ quasi sicuramente per $n \longrightarrow +\infty$ $\sharp$.

Stima di Hardy-Littlewood: $S_n - np = O(\sqrt{n \log n})$ *quasi sicuramente per* $n \longrightarrow +\infty$.

Cioè, $\exists n_0$ e $c > 0$ tali che $\forall n > n_0$, $|S_n - np| \leq c|\sqrt{n \log n}|$.

Dimostrazione: Faremo vedere che $S_n - np \leq \sqrt{n \log n}$ quasi sicuramente per $n \longrightarrow +\infty$. Il TGD ci dice che:

$$P\left(\frac{S_n}{n} \geq p + \epsilon\right) \leq e^{-nh_+(\epsilon)},$$

$\forall n \geq 1$, dove $h_+(\epsilon) = (p+\epsilon)\log\left(\frac{p+\epsilon}{p}\right) + (1-p-\epsilon)\log\left(\frac{1-p-\epsilon}{1-p}\right)$.

Ora, per $\epsilon \to 0$, $h_+(\epsilon) = \dfrac{\epsilon^2}{2p(1-p)} + O(\epsilon^2)$ (relazione(h')). Per vederlo, consideriamo la funzione estesa $v(\epsilon) = 0$ se $\epsilon = 0$, $v(\epsilon) = h_+(\epsilon)$ se $\epsilon \in (0, 1-p)$; notando che $v(0) = 0$, $v'(0) = 0$, $v''(\epsilon) = \dfrac{1}{p+\epsilon} - \dfrac{1}{1-p-\epsilon}$, e che $v'''(\epsilon) = \dfrac{1}{(p+\epsilon)^2} + \dfrac{1}{(1-p-\epsilon)^2}$, si può scomporre $v(\epsilon)$ in Taylor come segue:

$$v(\epsilon) = \frac{v''(0)\epsilon^2}{2!} + O(\epsilon^2),$$

dove

$$\frac{v''(0)\epsilon^2}{2!} = \frac{1-p-p}{2p(1-p)}\epsilon^2 = \frac{\epsilon^2}{2p(1-p)} - \frac{\epsilon^2}{1-p} = \frac{\epsilon^2}{2p(1-p)} + O(\epsilon^2).$$

Visto che tutto ciò vale per $\epsilon \longrightarrow 0$, quindi in un intorno di 0, si può considerare valida la relazione (h'). Si noti che $P\left(\frac{S_n}{n} \geq p + \epsilon\right) = P(S_n - np \geq n\epsilon)$, quindi, se $\epsilon = \sqrt{\dfrac{\log n}{n}}$, risulta che: $P(S_n - np \geq \sqrt{n \log n}) \leq \leq e^{-nh_+(\sqrt{\log n/n})}$.
Quindi:
$h_+\left(\sqrt{\dfrac{\log n}{n}}\right) = \dfrac{\log n}{2p(1-p)n} + o\left(\dfrac{1}{n}\right)$, poiché $O\left(\left(\dfrac{\log n}{n}\right)^{3/2}\right) = o\left(\dfrac{1}{n}\right)$ (infatti, se $a_n \in O\left(\left(\dfrac{\log n}{n}\right)\right)^{3/2}$, allora $\exists n_0 \in \mathbb{N}$ e $c > 0$ tali che $|a_n| \leq c\left(\dfrac{\log n}{n}\right)^{3/2}$ $\forall n \geq n_0$; pertanto, se $n \geq n_0$, allora $|\dfrac{1}{n^{-1}}a_n| = |na_n| \leq cn\dfrac{(\log n)^{3/2}}{n \cdot n^{1/2}} = c\dfrac{(\log n)^{3/2}}{n^{1/2}}$. Sia $m = (\log n)$; allora $n = e^m$ e $n^{1/2} = e^{m/2}$, dunque $\dfrac{(\log n)^{3/2}}{n^{1/2}} = \dfrac{m^{3/2}}{e^{m/2}} \longrightarrow 0$ per $m \longrightarrow +\infty$). Detto ciò:

$$\begin{aligned}\exp\left(-nh_+\left(\frac{\log n}{n}\right)\right) &= \exp\left(-\frac{1}{2p(1-p)}\log n + o(1)\right) \\ &= \exp\left(-\frac{\log n}{2p(1-p)}\right) \cdot \exp(o(1)) = n^{1/2p(1-p)} \cdot \exp(o(1)).\end{aligned}$$

Da qui, $\exp\left(-nh_+\left(\sqrt{\dfrac{\log n}{n}}\right)\right) \sim n^{1/2p(1-p)}$. Detto questo, poiché $\mathbb{R}$ è soluzione della disequazione $4x^2 - 4x + 1 \geq 0$, risulta che $p(1-p) \leq \dfrac{1}{4}$, da cui $2p(1-p) \leq \frac{1}{2}$, quindi $\frac{1}{2p(1-p)} \geq 2$, ragion per cui $-\frac{1}{2p(1-p)} \leq -2$, pertanto $n^{-1/2p(1-p)} \leq n^{-2}$. Di conseguenza, $\sum_{n\geq 1} n^{-1/2p(1-p)}$ è convergente, e quindi $\sum_{n\geq 1} P(S_n - np > \sqrt{n \log n}) < +\infty$ e dunque, concludendo, $P(S_n - np \leq \sqrt{n \log n}) = 1$ per n abbastanza grande $\sharp$.

Lemma 1: $\forall\ a, \delta > 0$, *e* $\forall n$ *sufficientemente grande,*

$$(\log n)^{-a^2(1+\delta)} < P(S_n - np > a \cdot \alpha(n)) < (\log n)^{-a^2(1-\delta)}.$$

Dimostrazione: Il TGD dice che:

$$P(R_n \geq a\alpha(n)) = P(S_n - np \geq a \cdot \alpha(n)) = P\left(\frac{S_n}{n} - p \geq \frac{a \cdot \alpha(n)}{n}\right) \leq$$

$$\leq exp\left(-nh_+\left(\frac{a\alpha(n)}{n}\right)\right).$$

Poiché $\frac{\alpha(n)}{n} \longrightarrow 0$ per $n \longrightarrow +\infty$, allora:

$$h_+\left(\frac{a\alpha(n)}{n}\right) = \frac{a^2}{2p(1-p)}\left(\frac{\alpha(n)}{n}\right)^2 + O\left(\left(\frac{\alpha(n)}{n}\right)^3\right);$$

quindi:

$$nh_+\left(\frac{a\alpha(n)}{n}\right) = a^2 \log\log n + O\left(\frac{\alpha(n)^3}{n^2}\right) \geq a^2(1-\delta)\log\log n$$

per n abbastanza grande.
Questo significa che:

$$P\left(\frac{S_n}{n} - p \geq a\alpha(n)\right) \leq \exp(-a^2(1-\delta)\log\log n) = (\log)^{-a^2(1-\delta)}.$$

Poiché $\sqrt{\log\log n} = o(n^{1/6})$, in virtù del Teorema delle Piccole Deviazioni, abbiamo che:

$$P\left(\frac{S_n}{n} - p \geq \frac{a\alpha(n)}{n}\right) = P\left(\frac{S_n}{n} - p \geq \sqrt{\frac{p(1-p)}{n}}a\sqrt{2n\log\log n}\right) \sim$$

$$\sim \frac{1}{\sqrt{2\pi}\cdot\sqrt{2\log\log n}}\exp(-a^2\log\log n) = \frac{1}{2a\sqrt{\pi\log\log n}}(\log n)^{-a^2}.$$

Poiché $\sqrt{\log\log n} = o((\log)^{a^2\delta})$, abbiamo che:

$$P\left(\frac{S_n}{n} - p \geq \frac{a\alpha(n)}{n}\right) \geq (\log n)^{-a^2(1+\delta)}$$

per n sufficientemente grande $\sharp$.

Disuguaglianza di Chebyshev: *Sia X una v.a. con media $\mu < +\infty$ e varianza $\sigma^2 < +\infty$ e sia $k > 0$. Allora, vale la seguente disuguaglianza:*

$$P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Dimostrazione: Poiché $(X-\mu)^2$ è una variabile aleatoria positiva, si può applicare su di essa la Disuguaglianza di Markov, per ogni $a > 0$:

$$P((X-\mu)^2 \leq a) \geq \frac{E[(X-\mu)^2]}{a};$$

ponendo $a = k^2$, si ottiene:

$$P((X-\mu)^2 \geq k^2) \leq \frac{E[(X-\mu)^2]}{k^2}.$$

Ma $(X-\mu)^2 \geq k^2$ sse $|X-\mu| \geq k^2$, peranto, si può scrivere:

$$P(|X-\mu| \geq k) \leq \frac{E[(X-\mu)^2]}{k^2}.$$

Inoltre, $E[(X-\mu)^2] = var(X) = \sigma^2$. Perciò, $P(|X-\mu| \geq k) \leq \dfrac{\sigma^2}{k^2}$ $\sharp$.

Lemma 2 (Disuguaglianza Massimale di Kolmogorov): *Sia $(Y_n)_{n\in\mathbb{N}}$ una successione di variabili aleatorie indipendenti e supponiamo che $E[Y_n] = 0$ e che $var(Y_n) = \sigma^2$. Definiamo $T_n = Y_1 + ... + Y_n$. Allora:*

$$P\left(\max_{1\leq k\leq n} T_k \geq b\right) \leq \frac{4}{3} P(T_n \geq b - 2\sigma\sqrt{n}).$$

Dimostrazione: Poiché le Y_n sono indipendenti, allora $var(T_n - T_k) = (n-k)\sigma^2$ per $1 \leq k \leq n$. La Disuguaglianza di Chebyshev ci dice che:

$$P(|T_n - T_k| \leq 2\sigma\sqrt{n}) \geq 1 - \frac{var(T_n - T_k)}{4\sigma^2 n} = 1 - \frac{n-k}{4n} \geq \frac{3}{4}.$$

Adesso, notiamo che

$$P\left(\max_{1\leq k\leq n} T_k \geq b\right) = \sum_{k=1}^{n} P(T_1 < b, ..., T_{k-1} < b, T_k \geq b) \leq$$

$$\leq \sum_{K=1}^{n} P(T_1 < b, ..., T_{k-1} < b, T_k \geq b) \frac{4}{3} P(|T_n - T_k| \leq 2\sigma\sqrt{n}) = (*)$$

Ora, per via dell'indipendenza fra gli eventi $(T_1 < b, ..., T_{k-1} < b, T_k \geq b)$ e $(|T_n - T_k| \leq 2\sigma\sqrt{n})$, vale che:

$$(*) = \frac{4}{3} \sum_{k=1}^{n} P(T_1 < b, ..., T_{k-1} < b, T_k \geq b, |T_n - T_k| \leq 2\sigma\sqrt{n}) \leq$$

$$\leq \frac{4}{3}\sum_{k=1}^{n} P(T_1 < b, ..., T_{k-1} < b, T_k \geq b, T_n \geq b - 2\sigma\sqrt{n}) \leq \frac{4}{3}P(T_n \geq b - 2\sigma\sqrt{n})\sharp.$$

Lemma 3: *Sia $(Y_n)_{n\in\mathbb{N}}$ una successione di variabili aleatorie identicamente distribuite ed indipendenti, con $E[Y_n] = 0$ e $var(Y_n) = \sigma^2 < +\infty$. Sia $T_n = \sum_{k=1}^{n} Y_n$. Allora:*

$$\epsilon^2 P\left(\max_{1\leq k\leq n} |T_k| > \epsilon\right) \leq n\sigma^2$$

per ogni $\epsilon > 0$.

Lemma 4: *Sia $(X_n, \mathfrak{F}_n)_{n\in\mathbb{N}}$ una sotto-martingala (ossia con $E[X_{n+1}|\mathfrak{F}_n] \geq X_n$ $\forall n \in \mathbb{N}$), tale che $X_n \geq 0$ qu.s. $\forall n \in \mathbb{N}$, e sia $\lambda > 0$. Allora, $\forall n \in \mathbb{N}$,*

$$\lambda \cdot P\left(\max_{1\leq k\leq n} X_n > \lambda\right) \leq E[X_n].$$

Dimostrazione della LLI:

Prima parte: Faremo vedere che $\limsup_{n\longrightarrow+\infty} \dfrac{S_n - np}{\alpha(n)} < 1 + \eta$ qu.s $\forall\eta > 0$.
Sia dunque $\eta > 0$, e sia $\gamma > 1$; consideriamo quindi l'applicazione : $k \in \mathbb{Z} \longmapsto n_k = \llcorner\gamma^k\lrcorner$. Facciamo vedere che:

$$\sum_{k\geq 1} P\left(\max_{n\leq n_k+1}(S_n - np) \geq (1+\eta)\alpha(n_k)\right) < +\infty.$$

Per il Lemma 2,

$$\sum_{k\geq 1} P\left(\max_{n\leq n_k+1}(S_n - np) \geq (1+\eta)\alpha(n_k)\right) \leq$$

$$\frac{4}{3}P(R_{n_k+1} \geq (1+\eta)\alpha(n_k) - 2\sqrt{n_{k+1}p(1-p)})$$

, dove $R_n = S_n - np$.
Osserviamo che $\sqrt{n_{k+1}} = o(\alpha(n_k))$, ragion per cui $2\sqrt{n_{k+1}p(1-p)} < \dfrac{1}{2}\eta\alpha(n_k)$ per k abbastanza grande. Detto questo:

$$P\left(\max_{n\leq n_k+1}(S_n - np) \geq (1+\eta)\alpha(n_k)\right) \leq \frac{4}{3}P(S_{n_{k+1}} - n_{k+1}p \geq (1+\frac{\eta}{2})\alpha(n_k)).$$

Ora, $\alpha(n_{k+1}) \sim \sqrt{\gamma}\alpha(n_k)$; scelgo quindi un γ tale che $1 + \dfrac{\eta}{2} > (1 + \dfrac{\eta}{4}\alpha(n_{k+1}))$.
Utilizzando il Lemma 1, si ottiene che:

$$P\left(\max_{n\leq n_k+1}(S_n - np) \geq (1+\eta)\alpha(n_k)\right) \leq \frac{4}{3}(\log n_{k+1})^{-(1+\frac{\eta}{4})}$$

per k molto grande. Notiamo adesso che $(\log n_{k+1})^{-(1+\frac{\eta}{4})} \sim (\log \gamma)^{-(1+\frac{\eta}{4})}k^{-(1+\frac{\eta}{4})}$, che è in generale il termine di una serie convergente, pertanto:

$$\sum_{k\geq 1} P\left(\max_{n\leq n_k+1} R_n \geq (1+\eta)\alpha(n_k)\right) < +\infty.$$

Ora: tutto ciò, implica che $\max_{n\leq n_k+1}(S_n - np) < (1+\eta)\alpha(n_k)$ qu.s. per k sufficientemente grande. Quindi, in particolare, $\max_{n_k\leq n\leq n_{k+1}}(S_n - np) < (1+\eta)\alpha(n_k)$ per k sufficientemente grande. Questo implica che, quasi sicuramente, $S_n - np < (1+\eta)\alpha(n_k)$, per $n > n_k$ e k abbastanza grande.

Parte 2: Mostriamo che, $\forall \eta > 0$, $\limsup_{n\longrightarrow +\infty} > 1 - \eta$ quasi ovunque. Per questa parte, ci serviremo del *Lemma di Borel-Cantelli:*

Lemma: *Sia* $(A_n)_{n\in\mathbb{N}}$ *una successione di eventi indipendenti su un certo spazio di probabilità. Se*

$$\sum_{n=1}^{+\infty} P(A_n) = +\infty,$$

allora $P(A) = 1$, *dove* $A = \bigcap_{k=1}^{+\infty}\bigcup_{n\geq k} A_n$.

Ora, sarà sufficiente mostrare che $\exists n_k$ tale che $R_{n_k} \geq (1-\eta)\alpha(n_k)$ quasi sicuramente per k abbastanza grande, e per farlo faremo vedere che per un $\gamma \in \mathbb{Z}$ opportuno, se $n_k = \gamma^k$, vale la seguente equazione, che chiameremo (D'):

$$\sum_{n\geq 1} P(R_{\gamma^n} - R_{\gamma^{n-1}} \geq (1-\frac{\eta}{2})\alpha(\gamma^n)) = +\infty,$$

e quindi anche che $R_{\gamma^{n-1}} \geq -\eta 2\alpha(\gamma^n)$ qu.o., per n sufficientemente grande. Si noti che $R_{\gamma^n} - R_{\gamma^{n-1}} = R_{\gamma^n - \gamma^{n-1}}$ in distribuzione; sarà pertanto sufficiente considerare la quantità $P(R_{\gamma^n-\gamma^{n-1}} \geq (1-\frac{\eta}{2})\alpha(\gamma^n))$. Si faccia innanzitutto caso al fatto che:

$$\frac{\alpha(\gamma^n - \gamma^{n-1})}{\alpha(\gamma^n)} \longrightarrow \sqrt{1-\frac{1}{\gamma}}$$

per $n \longrightarrow +\infty$. Scegliamo un $\gamma \in \mathbb{Z}$ tale che $\dfrac{1-\frac{\eta}{2}}{1-\frac{\eta}{4}} < \sqrt{1-\dfrac{1}{\gamma}}$. Allora, si può scegliere un n abbastanza grande da rendere vera la disuguaglianza: $\dfrac{1-\frac{\eta}{2}}{1-\frac{\eta}{4}} < \dfrac{\alpha(\gamma^n-\gamma^{n-1})}{\alpha(\gamma^n)}$, da cui $(1-\frac{\eta}{2})\alpha(\gamma^n) < (1-\frac{\eta}{4})\alpha(\gamma^n-\gamma^{n-1})$. Da qui, otteniamo:

$$P(R_{\gamma^n} - R_{\gamma^{n-1}} \geq (1-\frac{\eta}{2})\alpha(\gamma^n)) \geq P(R_{\gamma^n-\gamma^{n-1}} \geq (1-\frac{\eta}{4})\alpha(\gamma^n - \gamma^{n-1})).$$

Adesso, per il Lemma 1, prendendo un δ tale che $a = (1+\delta)^{-1} = (1-\frac{\eta}{4})$, risulta che:

$$P(R_{\gamma^n} - \gamma^{n-1} \geq (1-\frac{\eta}{4})\alpha(\gamma^n - \gamma^{n-1})) \geq (\log(\gamma^n - \gamma^{n-1}))^{-(1-\eta/4)} =$$

$$= (n \log \gamma + \log(1 + 1/\gamma))^{-(1-\eta/4)},$$

ed una serie con tali termini è sempre divergente, pertanto la (D') è dimostrata. Detto ciò, osserviamo adesso che $\alpha(\gamma^n) \sim \sqrt{\gamma}\alpha(\gamma^{n-1})$. Scegliamo ora un γ tale che $\eta\sqrt{\gamma} > 4$. Allora, $\eta\alpha(\gamma^n) \sim \eta\sqrt{\gamma\alpha(\gamma^{n-1})} > 4\alpha(\gamma^{n-1})$ per n sufficientemente grande. Di conseguenza, abbiamo che:

$$(R_{\gamma^{n-1}} \leq \frac{-\eta}{2}\alpha(\gamma^n)) \subseteq (-R_{\gamma^{n-1}} \geq 2\alpha(\gamma^{n-1})).$$

Ora, servendosi dell' equazione (D'), possiamo affermare che $R_{\gamma^{n-1}} < 2\alpha(\gamma^{n-1})$ quasi sicuramente per n sufficientemente grande. Adesso, $R_{\gamma^n} - R_{\gamma^{n-1}} \geq (1 - \frac{\eta}{2})\alpha(\gamma^n)$ per $n >> 1$; aggiungendoci la disuguaglianza $R_{\gamma^{n-1}} \geq -\frac{\eta}{2}$, otteniamo che $R_{\gamma^n} > (1-\eta)\alpha(\gamma^n)$ per $n >> 1$. Questo è sufficiente per dimostrare che $\liminf_{n \longrightarrow +\infty} \dfrac{S_n - np}{\alpha(n)} > 1 - \eta$ quasi sicuramente, che è sufficiente per dimostrare la seconda parte della LLI di Khinchin $\sharp$.

Che interpretazione si può dare alla LLI? Innanzitutto, ha in comune con la LGN il fatto di analizzare un processo stocastico sui singoli cammini, cioè, data, una sequenza di n ripetizioni dello stesso fenomeno aleatorio, si può notare come, se n è molto grande, tale sequenza tenda a stabilizzarsi su una data situazione, nonostante l'indipendenza dei singoli tentativi effettuati: tali tentativi sono accomunati dal comportamento globale del processo che si sta prendendo in considerazione. Si può far caso poi al fatto che la LLI di Khinchin ci dice come un processo (in generale si può prenderne in esame uno simile al lancio ripetuto di una monetina), oscilli in continuazione fra due situazioni estreme: da una parte, per n molto grande, $S_n - np$ sarà destinato ad essere molto vicino al valore $\alpha(n) = \sqrt{2np(1-p)\log\log n}$, mentre per un n ancora più grande, il valore $S_n - np$ sarà destinato ad avvicinarsi all'estremo opposto, ossia $-\alpha(n)$, e viceversa. Questo significa che se si simula una successione di femoneni aleatori, con $X_n \sim B_p$, e si disegna il grafico dei risultati con $x = n$ ed $y = S_n - np$, allora tali risultati saranno sempre circoscritti al grafico G della curva di equazione $y^2 = 2px(1-p)\log\log x$, $x \in (1, +\infty)$ (se $n = 1$, allora $S_n - np = 1 - p$ oppure $S_n - np = -p$), i massimi ed i minimi relativi della successione $S_n - np$ saranno destinati ad avvicinarsi in maniera asintotica a G con il progressivo aumento del valore di n. Prendiamo ora in considerazione il TLC: che cosa differenzia questo dalla LGN forte e dalla LLI? Il TLC analizza un processo aleatorio in base a quello che è il suo comportamento in media: viene stimata la probabilità di trovarsi in una certa gamma di valori, in relazione alla situazione di stabilià ideale descritta dalla LGN forte, quindi si calcola di fatto una misura del peso di un dato evento specifico.

Ciò detto, come visto nel capitolo sul TLC, quest'ultimo ci dà una quantità di informazioni maggiore rispetto alla LGN forte; ebbene, la LLI ci fornisce un grado di precisione ancora maggiore, ponendosi esattamente nel mezzo: da una parte, sappiamo che $\lim\limits_{n \longrightarrow +\infty} \dfrac{S_n - np}{n} = 0$; qui il problema risiede nel fatto che il fattore n è più forte del fattore $S_n - np$, quindi, di fatto, il valore n, considerato come funzione che ad ogni valore restituisce il valore stesso è uno strumento troppo potente, pertanto c'è una condensazione eccessiva di informazioni. D'altro canto, il TLC ci dice che $\lim_{n \longrightarrow +\infty} \frac{S_n - np}{\sqrt{np(1-p)}} = Z \sim N(0,1)$: questo implica che per n molto grande circa il 68 per cento dei singoli tentativi della sequenza di lunghezza n, soddisfa $\left| \frac{S_n - np}{\sqrt{np(1-p)}} \right| \leq 1$, mentre il 95 per cento di essi, invece, soddisfa $\left| \frac{S_n - np}{\sqrt{np(1-p)}} \right| \leq 2$. Dunque, si ottiene l'effetto opposto, abbiamo che il denominatore $\sqrt{n}$ è troppo debole, cioè non condensa abbastanza informazioni e ne lascia disperdere una quantità eccessiva; si può inoltre dimostrare che $\liminf_{n \longrightarrow +\infty} \frac{S_n - np}{\sqrt{np(1-p)}} = -\infty$ e che $\limsup_{n \longrightarrow +\infty} \frac{S_n - np}{\sqrt{np(1-p)}} = +\infty$. Dunque, ci si può aspettare che fra $\sqrt{n}$ ed n ci siano valori che non condensino troppe informazioni, né ne lascino disperdere in quantità eccessiva: la LLI di Khinchin ci dice che $\sqrt{np(1-p) \log \log n}$ è uno di questi valori.

Bibliografia

[EL] Ermanno Lanconelli, *Lezioni di analisi matematica 2*, vol.2, Pitagora, 2001.

[JG] Jean Franois Le Gall, *Brownian Motion, Martingales, and Stochastic Calculus*, Springer Verlag, 2016.

[JW] John B. Walsh, *Notes on Elementary Martingale Theory*, notes, **https://www.math.ubc.ca/ walsh/marts.pdf**.

[HK] H. Krieger, *Proof of Central Limit Theorem*, Harvey Mudd College, **https://www.math.hmc.edu/ krieger/m157cltproof.pdf**, 2005.

[JT] Jay Taylor, *Lectures for STP 421: Probability Theory*, Arizona State University, lectures, **https://math.la.asu.edu/ jtaylor/teaching/Spring2017/STP421/lectures/probability.pdf**, 2016.

[DL] Don McLeish, *STAT 901:PROBABILITY*, notes, **http://sas.uwaterloo.ca/ dlmcleis/s901/s9012005.pdf**, 2005.

[MW] Matthias Winkel, *The Strong Law of Large Numbers*, lectures, **http://www.stats.ox.ac.uk/ winkel/bs3a07l9-10.pdf**

[MW] Michael Woodroofe, *The Martingale Central Limit Theorem*, notes, **http://dept.stat.lsa.umich.edu/ michaelw/STAT621W03/mrtglclt.pdf**, 2003.

[SD] Steven R. Dunbar, **Topics in Probability Theory and Stochastic Processes Steven R. Dunbar** , *Large Deviations*, University of Nebraska-Lincoln, lessons **https://www.math.unl.edu/ sdunbar1/ProbabilityTheory/Lessons/BernoulliTrials/LargeDeviations /largedeviations.pdf**, 2018.

[SD] Steven R. Dunbar, **Topics in Probability Theory and Stochastic Processes Steven R. Dunbar** , *The Moderate Deviations Result*, University of Nebraska-Lincoln, lessons, **https://www.math.unl.edu/ sdunbar1/ProbabilityTheory/Lessons/BernoulliTrials/ModerateDeviations /moderatedeviations.pdf**, 2012.

[SD] Steven R. Dunbar, **Topics in Probability Theory and Stochastic Processes Steven R. Dunbar** , *Law of the Iterated Logarithm*, **https://www.math.unl.edu/ sdunbar1/ProbabilityTheory/Lessons/BernoulliTrials/IteratedLogarithm /iteratedlogarithm.pdf**, 2018.