Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA “TULLIO LEVI-CIVITA”

Corso di Laurea in Matematica

Modelli stocastici per la legge di Benford

Relatore:
Prof. Marco Formentin

Laureanda: Alice Covelli
Matricola: 2033937

Anno Accademico 2023/2024

13 Dicembre 2024

Indice

Introduzione

La legge di Benford (BL), anche conosciuta come legge della prima cifra o legge della cifra significativa, è una delle più sorprendenti regolarità statistiche che si incontrano in insiemi di dati raccolti in diversi contesti. Il primo riferimento noto alla distribuzione logaritmica delle cifre significative risale al 1881: in questo anno l'astronomo americano Simon Newcomb pubblicò nell'American Journal of Mathematics un articolo intitolato "Note on the Frequency of Use of the Different Digits in Natural Numbers" nel quale affermò che in ogni lista di numeri ricavati da un insieme arbitrario di dati, la maggior parte di essi tende a iniziare per 1 e, al procedere delle cifre, corrisponde una diminuizione della frequenza con cui queste appaiono come first-digit.
Il modo in cui Newcomb arrivò alla formulazione di tale principio risulta bizzarro: egli si rese conto che i bordi delle prime pagine delle tavole logaritmiche (considerando le tavole aventi 1 come prima cifra) erano maggiormente usurati rispetto a quelli delle ultime, presumibilmente poiché sfogliate, di conseguenza utilizzate, più frequentemente. La sua scoperta però non possedeva il supporto di osservazioni empiriche, pertanto non attirò la dovuta attenzione e finì con l'essere dimenticata.
Quasi mezzo secolo più tardi, il fisico Frank Benford, ignaro del lavoro di Newcomb, riscoprì la legge ma, a differenza del suo predecessore, fornì osservazioni empiriche di essa, ottenendo conferme dalle osservazioni dirette e dalle riflessioni compiute su di esse. Egli supportò la sua scoperta raccogliendo una collezione di 20.000 dati numerici, relativi a 21 elementi di natura molto variegata (da bacini idrografici di 335 fiumi a calori specifici di 1.389 composti chimici, passando per le statistiche di baseball dell'American League) misurando la frequenza delle cifre da 1 a 9, compresi, escludendo dall'analisi la cifra 0. Nella Tabella troviamo alcuni dei dati presenti nell' articolo di Benford.
Nel concreto, la legge di Benford afferma che in molte raccolte di dati, siano essi per esempio dati reali o combinazioni di essi, le cifre significative iniziali, differentemente da come ci si potrebbe aspettare, non sono distribuite uni-

Titolo	1	2	3	4	5	6	7	8	9	N° campioni
Fiumi	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Popolazione	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
Costanti	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
Giornali	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
Calori Specifici	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
Pressione	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
Indirizzi	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
Tasso mortalità	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418

Tabella 1: Percentuale di volte in cui i numeri naturali da 1 a 9 appaiono come first-digit nei numeri.

formemente, bensì sono fortemente sbilanciate verso le cifre più piccole e, più precisamente, seguono una particolare distribuzione logaritmica. Nella sua più comune formulazione, detta "il caso speciale della prima cifra decimale significativa"(cioè la prima diversa da zero), la legge di Benford afferma che

$$\text{Prob}\,(D_1 = d) = \log\left(1 + \frac{1}{d}\right) \qquad \forall\ d = 1, 2, \ldots, 9 \tag{0.1}$$

dove D_1 indica la prima cifra decimale significativa.
La legge di Benford trova impiego in molti ambiti ed è pertanto importante nominare alcune tra le sue molte applicazioni.
Quella per cui la legge di Benford è maggiormente nota interessa il campo della revisione forense, in particolare nella rilevazione statistica della frode contabile, dove per "frode" si intende sia la fabbricazione dei dati sia la loro falsificazione (manipolazione e/o alterazione). L'idea di base è la seguente: se per alcune classi di set di dati finanziari veritieri si è osservato che seguono strettamente una distribuzione di Benford, allora a volte set di dati fabbricati o falsificati possono essere identificati semplicemente cercando deviazioni nelle frequenze delle singole cifre significative.
La ragione principale che ha portato a seguire questa linea di ricerca è stata la scoperta di Nigrini [10, 11] del fatto che alcuni dati fiscali certificati dall'IRS (Internal Revenue Service, agenzia governativa deputata alla riscossione dei tributi all'interno del sistema tributario degli Stati Uniti d'America) seguono la legge di Benford, mentre i dati fraudolenti spesso non lo fanno. Di conseguenza, se ci si aspetta che i dati autentici seguano approssimativamente tale legge, un cattivo adattamento ad essa solleva dei sospetti: a quel punto si procede applicando test indipendenti (non basati su Benford) o attraverso monitoraggi.

Il primo successo di questo metodo è stato riportato nel 1995 dall'Ufficio del Procuratore Distrettuale di Brooklyn a New York, dove il principale investigatore finanziario ha utilizzato un semplice test di adattamento alla legge della prima cifra per identificare e perseguire con successo sette aziende per furto [4].
Una domanda più intrigante è se la legge di Benford possa essere utilizzata per rilevare segnali in contrasto con il rumore di fondo, ad esempio, nei dati delle serie temporali. Qui entra in gioco il lavoro di M. Sambridge, H. Tkalčić e A. Jackson [14], i quali si sono posti il seguente quesito: può un terremoto essere rilevato semplicemente osservando le frequenze delle prime cifre significative dei conteggi degli spostamenti del terreno registrati da un sismometro? Per rispondere a questa domanda, hanno confrontato le distribuzioni delle prime cifre significative previste con quelle rilevate all'interno di una finestra temporale mobile di 200 secondi, per una durata di 40 minuti, centrata sugli arrivi delle prime onde PKP del terremoto. Le onde PKP, o onde di compressione primaria, sono tali per cui la terra si comprime e si espande nella stessa direzione in cui l'onda si propaga e sono molto utili per gli scienziati che studiano la struttura interna della Terra, poichè il loro percorso e la loro velocità dipendono dalla composizione e dallo stato fisico dei materiali attraversati. Una misura di "bontà" dell'adattamento alle previsioni della legge di Benford (BL) è stata calcolata per ogni finestra utilizzando

$$\phi = \left[1 - \left(\sum_{D=1}^{9} \frac{(n_D - nP_D)^2}{nP_D}\right)^{1/2}\right] \times 100$$

dove n_D è il numero di dati osservati aventi prima cifra significativa D, P_D è la proporzione dei dati aventi prima cifra significativa D prevista da (0.1) e n è il numero totale di dati. Dai dati raccolti hanno osservato che le prime cifre significative del rumore che precede l'arrivo del terremoto non seguono la legge di Benford (ϕ è minore di 0) ma, non appena la finestra mobile incontra le onde sismiche, ϕ inizia ad aumentare. L'adattamento continua ad aumentare costantemente man mano che nuovi segnali sismici vengono inclusi nella finestra temporale, il che dimostra chiaramente che la presenza dei terremoti può essere rilevata solo dalle informazioni sulle cifre, cioè senza mai vedere i dettagli di un sismogramma (grafico prodotto dal sismografo). Grazie a quanto scoperto, utilizzando solo le informazioni relative alla prima cifra significativa, sono inoltre riusciti a rilevare un terremoto locale vicino a Canberra, senza la necessità di disporre dei dettagli completi del sismogramma.
Ad ogni modo i test di adattamento alla legge di Benford hanno altri numerosi impieghi, come il valutare i risultati previsti da un modello matematico

o la qualità dell'output di uno strumento di rilevamento o, ancora, analizzare vari tipi di errori computazionali nell'ambito dell'aritmetica in virgola mobile.
La legge di Benford appare in un'ampia varietà di statistiche, tra cui modelli stocastici, come quello del moto browniano geometrico (un processo stocastico positivo a tempo continuo spesso utilizzato per modellizzare alcuni fenomeni dei mercati finanziari), statistiche d'ordine, modelli bayesiani e processi di Lévy (processi stocastici con incrementi stazionari ed indipendenti, rappresentano il moto di un punto i cui movimenti successivi siano indipendenti e siano identicamente distribuiti su intervalli di tempo della stessa lunghezza).
Essa emerge inoltre frequentemente anche in funzioni a valori reali, come quelle che sorgono come soluzioni di problemi ai valori iniziali per equazioni differenziali, tuttavia tale argomento non verrà trattato in questa tesi, la quale si concentrerà invece sull'applicazione della legge di Benford ai processi stocastici.
I riferimenti principali di questa tesi sono i risultati contenuti in "An Introduction to Benford's Law"[2] e "The Mathematics of Benford's Law - A Primer"[3] di A. Berger e T. P. Hill.
Contenuto dei Capitoli. Nel *Capitolo 1* getteremo le basi per affrontare la legge di Benford in termini probabilistici, introducendo i concetti di funzione e σ-algebra dei significandi, distribuzione di Benford e variabile aleatoria Benford. Si esaminerà poi come molte delle distribuzioni più comuni (come quella esponenziale o gamma) non seguano la legge di Benford, ma che alcune distribuzioni, come quella Pareto, possano avvicinarsi ad essa sotto particolari condizioni. Infine si vedranno alcuni risultati importanti che legano la legge di Benford alla distribuzione uniforme modulo uno, i quali saranno centrali per la trattazione degli argomenti del Capitolo 2.
Nel *Capitolo 2* i temi di interesse saranno principalmente due:

- evidenziare come la legge di Benford sia fortemente attrattiva, studiando la convergenza in distribuzione e la convergenza quasi certa alla legge di Benford di successioni i.i.d. di potenze, prodotti e somme di variabili aleatorie;

- giustificare perchè nello studio di Benford molte delle sue tabelle non si conformassero alla legge logaritmica, al contrario invece dell'unione di tali tabelle; ciò verrà fatto attraverso il Teorema 2.19.

Capitolo 1

Distribuzione di Benford e variabili aleatorie

In questo capitolo verranno gettate le fondamenta per l'analisi della proprietà di Benford nelle distribuzioni di probabilità e nelle variabili aleatorie.
Prima di iniziare ad entrare nel merito però, è bene introdurre la funzione dei significandi e la σ-algebra dei significandi. Da qui in poi si userà log per indicare il logaritmo in base 10, mentre eventuali altre basi verranno specificate.

Definizione 1.1. *Per ogni $x \in \mathbb{R}$, $x \neq 0$, la prima cifra (decimale) significativa di x, che indicheremo con $D_1(x)$, è l'unico $j \in \{1, 2, \dots, 9\}$ che soddisfa $10^k j \leq |x| < 10^k(j+1)$ per un qualche (necessariamente unico) $k \in \mathbb{Z}$.*
Per ogni $m \geq 2$, $m \in \mathbb{N}$, la m-esima cifra (decimale) significativa di x, denotata con $D_m(x)$, è definita per induzione come l'unico $j \in \{1, 2, \dots, 9\}$ tale che

$$10^k \left(\sum_{i=1}^{m-1} D_i(x) 10^{m-i} + j \right) \leq |x| < 10^k \left(\sum_{i=1}^{m-1} D_i(x) 10^{m-i} + j + 1 \right)$$

per un qualche (necessariamente unico) $k \in \mathbb{Z}$. Imponiamo che $D_m(0) := 0$.

Definizione 1.2. *La funzione dei significandi (decimali) $S\colon \mathbb{R} \to [1, 10)$ è definita come segue:*

- *se $x \neq 0$ allora $S(x) = t$ dove t è l'unico numero in [1,10) tale che $|x| = 10^k t$ per un qualche (necessariamente unico) $k \in \mathbb{Z}$;*
- *se $x = 0$ allora $S(0) := 0$.*

Esplicitamente, la funzione S è data da

$$S(x) = 10^{\log|x| - \lfloor \log|x| \rfloor} \text{ per ogni } x \neq 0$$

dove con $\lfloor x \rfloor$ si denota il più grande intero minore o uguale al numero reale x.

La prossima proposizione, che segue immediatamente dalla Definizione 1.1 e dalla Definizione 1.2, mette in luce come il significando determini in modo univoco tutte le cifre significative, e viceversa, esprimendo dapprima il significando di un numero come funzione esplicita delle cifre significative di quel numero, e poi le cifre significative come funzione della parte significativa.

Proposizione 1.1. *Sia $x \in \mathbb{R}$. Allora:*

i. $S(x) = \sum_{m \in \mathbb{N}} 10^{1-m} D_m(x)$;

ii. $D_m(x) = \lfloor 10^{m-1} S(x) \rfloor - 10 \lfloor 10^{m-2} S(x) \rfloor$ *per ogni* $m \in \mathbb{N}$.

Definizione 1.3. *La σ-algebra dei significandi è la σ-algebra su $\mathbb{R}^+$ generata dalla funzione dei significandi S, ossia $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S)$.*

Lemma 1.1. *Per ogni funzione $f\colon \mathbb{R}^+ \to \mathbb{R}$ le seguenti affermazioni sono equivalenti:*

i. *f è completamente determinata da S, cioè esiste una funzione $\varphi\colon [1, 10) \to \mathbb{R}$ con $\sigma(\varphi) \subset \mathcal{B}[1, 10)$ tale che $f(x) = \varphi(S(x))$ per ogni $x \in \mathbb{R}^+$;*

ii. $\sigma(f) \subset \mathcal{S}$.

Dimostrazione. Supponiamo che valga (i), vogliamo dimostrare che allora vale (ii).
Sia $J \subset R$ intervallo qualsiasi. Allora $B = \varphi^{-1}(J) \in \mathcal{B}$ e, poiché per ipotesi vale (i), $f^{-1}(J) = S(\varphi^{-1}(J)) = S^{-1}(B) \in \mathcal{B}$, ciò dimostra che $\sigma(f) \in \mathcal{S}$.
Viceversa, supponiamo che valga (ii), mostriamo che vale anche (i). Poiché per ipotesi (ii) vale $\sigma(f) \subset \mathcal{S}$, allora diciamo che $f(10x) = f(x)$ per ogni $x > 0$. Supponiamo infatti per assurdo che $f(x_0) < f(10x_0)$ per un qualche $x_0 > 0$, poniamo

$$A = f^{-1}\left([f(x_0) - 1, \frac{1}{2}(f(x_0) + f(10x_0)] \right) \in \sigma(f) \subset \mathcal{S}$$

e si osservi che $x_0 \in A$, mentre $10x_0 \notin A$. Dato che $A = S^{-1}(B)$ per un qualche $B \in \mathcal{B}$, ciò porta ad una contraddizione: stiamo affermando che

$S(x_0) \in B$, dunque $S(x_0) = S(10x_0)$, pertanto $S(10x_0) \in B$, ma $S(10x_0) \notin B$ per quanto appena osservato. Dunque $f(10x) = f(x)$ per ogni $x > 0$, e per induzione $f(10^k x) = f(x)$ per ogni $k \in \mathbb{Z}$.
Dato $t \in [1, 10)$, si scelga un $x > 0$ con $S(x) = t$ e definiamo $\varphi(t) = f(x)$. Poiché qualsiasi due scelte di x differiscono di un fattore 10^k per un certo $k \in \mathbb{Z}$, $\varphi\colon [1, 10) \to \mathbb{R}$ è ben definita e $f(x) = \varphi(S(x))$ per ogni $x > 0$. Per di più, ogni intervallo $J \subset \mathbb{R}$ e $t \in [1, 10)$, $\varphi(t) \in J$ se e solo se
$t \in \bigcup_{k\in\mathbb{Z}} 10^k f^{-1}(J)$. Per ipotesi, quest' ultimo insieme appartiene ad $\mathcal{S}$, il che a suo volta dimostra che $\sigma(\varphi) \subset \mathcal{B}[1, 10)$.

□

Il Lemma 1.1 afferma quindi che la σ-algebra $\mathcal{S}$ è la famiglia di tutti gli eventi $A \subset \mathbb{R}^+$ che possono essere completamente descritti in termini dei loro significandi. Il seguente teorema dà invece un'utile caratterizzazione dei membri di tale famiglia $\mathcal{S}$. Prima di enunciarlo, introduciamo la seguente notazione: per $x \in \mathbb{R}$ e per ogni insieme $C \subset \mathbb{R}$ poniamo

$$xC = \{xc : c \in C\}.$$

Teorema 1.1. *Per ogni $A \in \mathcal{S}$,*

$$A = \bigcup_{k\in\mathbb{Z}} 10^k S(A)$$

dove $S(A) = \{S(x) : x \in A\} \subset [1, 10)$. Inoltre,

$$\mathcal{S} = \mathbb{R} \cap \sigma(D_1, D_2, D_3 \dots) = \left\{ \bigcup_{k\in Z} 10^k B : B \in \mathcal{B}[1, 10) \right\}.$$

Dimostrazione. Per definizione

$$\begin{aligned} \mathcal{S} &= \mathbb{R}^+ \cap \sigma(S) \\ &= \mathbb{R}^+ \cap \{S^{-1}(B) : B \in \mathcal{B}\} \\ &= \mathbb{R}^+ \cap \{S^{-1}(B) : B \in \mathcal{B}[1, 10)\}. \end{aligned}$$

Dunque, dato un qualsiasi $A \in \mathcal{S}$, esiste $B \in \mathcal{B}[1, 10)$ con

$$A = \mathbb{R}^+ \cap S^{-1}(B) = \bigcup_{k\in Z} 10^k B$$

e poiché $B = S(A)$ segue quanto voluto.
Per provare il secondo asserto si osservi prima che per (i) della Proposizione 1.1 la funzione S è completamente determinata dalle cifre significative

$D_1, D_2, D_3 \ldots$, dunque $\sigma(S) \subset \sigma(D_1, D_2, D_3 \ldots)$ e, pertanto, $\mathcal{S} = \mathbb{R}^+ \cap \sigma(S) \subset \mathbb{R}^+ \cap \sigma(D_1, D_2, D_3 \ldots)$. D'altra parte, (ii) della Proposizione 1.1 mi dice che ogni D_m è determinato da S; di conseguenza $\sigma(D_m) \subset \sigma(S)$ per ogni $m \in \mathbb{N}$ e ciò mostra pertanto che $\sigma(D_1, D_2, D_3 \ldots) \subset \sigma(S)$. Per dimostrare la rimanente parte dell'uguaglianza, si osservi che per ogni $A \in \mathcal{S}$, $S(A) \in \mathcal{B}[1, 10)$ e dunque $A = \bigcup_{k \in Z} 10^k B$ con $B = S(A)$ per quanto mostrato precedentemente. Infine, ogni insieme della forma $\bigcup_{k \in Z} 10^k B = \mathbb{R}^+ \cap S^{-1}(B)$ con $B \in \mathcal{B}[1, 10)$ appartiene chiaramente a $\mathcal{S}$.

□

Possiamo ora entrare nel vivo di questo capitolo, iniziando con il dare la definzione di misura di probabilità Benford, di successione Benford e di variabile aleatoria Benford.

Definizione 1.4. *Una misura di probabilità di Borel P su $\mathbb{R}$ si dice Benford se*

$$P\left(\{x \in \mathbb{R} : S(x) \leq t\}\right) = P\left(S^{-1}\left(\{0\} \bigcup [1, t]\right)\right) = \log t \qquad \forall t \in [1, 10).$$

Una successione (x_n) di numeri reali è detta Benford se

$$\lim_{N \to \infty} \frac{\#\{1 \leq n \leq N : S(x_n) \leq t\}}{N} = \log t \quad \forall t \in [1, 10).$$

Una variabile aleatoria X in uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ è Benford se P_X è Benford , ovvero se

$$F_{S(X)}(t) = \mathbb{P}(S(X) \leq t) = P_X\left(\{x \in \mathbb{R} : S(x) \leq t\}\right) = \log t \quad \forall t \in [1, 10)$$

o, equivalentemente, se S(X) è una variabile aleatoria assolutamente continua con densità $f_{S(X)}(t) = t^{-1} \log e$. In termini delle sue cifre significative, X è Benford se e solo se

$$\mathbb{P}\big(D_j(X) = d_j \ \ \text{per } j = 1, 2, \ldots, m\big) = \log\left(1 + \left(\sum_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right),$$

$\forall m \in \mathbb{N}, \forall d_1 \in \{0, 1, \ldots, 9\}$ *e* $\forall d_j \in \{0, 1, \ldots, 9\}, j \geq 2$.

Dalla Definizione 1.4 segue immediatamente che per ogni variabile aleatoria Benford X, le cifre significative D_1(X), D_2(X), ... di X sono asintoticamente indipendenti e uniformemente distribuite , ovvero

$$\lim_{n \to \infty} \mathbb{P}\left(D_{n+1}\left(X\right) = d_1, D_{n+2}\left(X\right) = d_2, \ldots, D_{n+m}\left(X\right) = d_m\right) = 10^{-m}$$

$\forall m \in \mathbb{N}$ *e* $\forall d_1, d_2, \ldots, d_m \in \{0, 1, \ldots, 9\}$, ma ciò è valido per ogni variabile aleatoria con densità, pertanto in tal senso non c'è nulla di speciale riguardo alla distribuzione di Benford.

Definizione 1.5. *La distribuzione di Benford* $\mathbb{B}$ *è l'unica misura di probabilità su* $(\mathbb{R}^+, \mathcal{S})$ *tale che*

$$\mathbb{B}(S \leq t) = \mathbb{B}\left(\bigcup_{k\in\mathbb{Z}} 10^k[1,t]\right) = \log t \quad \forall t \in [1, 10)$$

o, equivalentemente, $\forall m \in \mathbb{N}, \forall d_1 \in \{0, 1, \ldots, 9\}$ *e* $\forall d_j \in \{0, 1, \ldots, 9\},\ j \geq 2$

$$\mathbb{B}(D_j = d_j \text{ per } j = 1, 2, \ldots, m) = \log\left(1 + \left(\sum_{j=1}^{m} 10^{m-j} d_j\right)^{-1}\right).$$

Data una misura di probabilità qualsiasi P su $(\mathbb{R}, \mathcal{B})$, denoteremo con $|P|$ la probabilità individuata da

$$|P|(B) = P\left(\{x \in \mathbb{R} : |x| \in B\}\right) \quad \forall B \in \mathcal{B}.$$

Risulta evidente che $|P|$ è concentrata in $[0, \infty)$, ossia $|P|\left([0, +\infty)\right) = 1$ e

$$F_{|P|}(x) = \begin{cases} 0 & \text{se } x < 0 \\ F_P(x) - F_P(-x) + P(\{-x\}) & \text{se } x \geq 0 \end{cases}$$

In particolare quindi, se P è una misura assolutamente continua, lo è anche $|P|$ e in tal caso la sua densità sarà $f_P(x) + f_P(-x)$ $\forall x \leq 0$, dove f_P rappresenta la densità di P.

Teorema 1.2. *Sia X una variabile aleatoria. Le seguenti affermazioni sono equivalenti*

i. *X è Benford;*

ii. $P(|X| \in A) = \mathbb{B}(A)$ $\forall A \in \mathcal{S}$;

iii. $E\left[S(X)^n\right] = (10^n - 1)n^{-1} \log e$ $\forall n \in \mathbb{N}$.

Osservazione 1.1. *La distribuzione di Benford* $\mathbb{B}$ *è una distribuzione di probabilità sulle cifre significative, o sui significandi, dei dati sottostanti e non sui dati grezzi stessi, ovvero* $\mathbb{B}$ *è una misura di probabilità sulla famiglia di insiemi definiti dai significandi base 10, ovvero su* $(\mathbb{R}^+, \mathcal{S})$*, ma non su* $(\mathbb{R}^+, \mathcal{B}^+)$ *o su* $(\mathbb{R}, \mathcal{B})$*. Di conseguenza, per esempio, la probabilità* $\mathbb{B}(\{1\})$ *non è definita: l'insieme costituito da un singolo elemento* $\{1\}$ *non può essere definito in termini di cifre significative o significandi da solo, e, pertanto, non appartiene al dominio di* $\mathbb{B}$.

Nessuna delle distribuzioni di probabilità classiche o variabili casuali comuni, come la distribuzioni uniforme, esponenziale, Pareto, normale, beta, binomiale o gamma, è esattamente Benford. D'altra parte, alcune famiglie di distribuzioni parametriche standard possono avvicinarsi arbitrariamente alla legge di Benford per alcuni valori dei parametri, mentre altre famiglie di distribuzioni non seguono la legge di Benford per tutti i valori dei parametri.
I prossimi due teoremi illustrano questo fenomeno prendendo in esame rispettivamente le famiglie delle distribuzioni di Pareto e uniforme positiva.

Teorema 1.3. *Sia X_α una variabile aleatoria Pareto di parametro $\alpha > 0$, ossia $\mathbb{P}(X_\alpha > x) = x^{-\alpha} \quad \forall x \geq 1$. Allora $\max_{1 \leq t < 10} |F_{S(X_\alpha)}(t) - \log t|$ converge in modo monotono a 0 per $\alpha \searrow 0$.*

Dimostrazione. Si osservi innanzitutto che per ogni $1 < t < 10$, la funzione $h_t \colon \mathbb{R} \to \mathbb{R}$, definita come

$$h_t(x) = \begin{cases} \frac{t^x - 1}{10^x - 1} & \text{se } x \neq 0 \\ \log(t) & \text{se } x = 0 \end{cases}$$

i. è continua: chiaramente è continua per $x \neq 0$ poiché composizione di funzioni continue, resta da verificare che sia continua in $x = 0$, ovvero dobbiamo mostrare che

$$\lim_{x \to 0^-} h_t(x) = \lim_{x \to 0^+} h_t(x) = h_t(0).$$

Pertanto

$$\lim_{x \to 0^-} h_t(x) = \lim_{x \to 0^-} \frac{t^x - 1}{10^x - 1} \overset{\text{Teorema di de l'Hopital}}{=} \lim_{x \to 0^-} \frac{t^x \ln(t)}{10^x \ln(10)} = \log t$$

$$\lim_{x \to 0^+} h_t(x) = \lim_{x \to 0^+} \frac{t^x - 1}{10^x - 1} \overset{\text{Teorema di de l'Hopital}}{=} \lim_{x \to 0^+} \frac{t^x \ln(t)}{10^x \ln(10)} = \log t$$

e $h_t(0) = \log t$ per definizione;

ii. è strettamente decrescente;

iii.
$$\lim_{x \to -\infty} h_t(x) = \lim_{x \to -\infty} \frac{t^x - 1}{10^x - 1} = 1;$$

iv.
$$\lim_{x \to \infty} h_t(x) = \lim_{x \to \infty} \frac{t^x - 1}{10^x - 1} \sim \lim_{x \to \infty} \left(\frac{t}{10}\right)^x = 0.$$

Si ha inoltre che per $1 < t < 10$ e $\alpha > 0$,

$$\begin{aligned}
F_{S(X_\alpha)}(t) &= \mathbb{P}\left(X_\alpha \in \bigcup_{k=0}^{\infty}[10^k, 10^k t]\right) \\
&= \sum_{k=0}^{\infty}\left(10^{-\alpha k} - 10^{-\alpha k} t^{-\alpha}\right) \\
&= \frac{1}{1-10^{-\alpha}} - \frac{t^{-\alpha}}{1-10^{-\alpha}} \\
&= \frac{t^{-\alpha}-1}{10^{-\alpha}-1} \\
&= h_t(-\alpha)
\end{aligned}$$

Di conseguenza, poiché $h_t(0) = \log(t)$, $\alpha > 0$ e h_t è strettamente decrescente, si ha che $h_t(-\alpha) > h_t(0)$, da cui segue che

$$F_{S(X_\alpha)}(t) - \log(t) = h_t(-\alpha) - h_t(0) > 0$$

è strettamente decrescente per $\alpha \searrow 0$ e pertanto lo sarà anche $\max_{1\leq t<10}\left|F_{S(X_\alpha)}(t) - \log t\right|$. Inoltre tende a 0 per $\alpha \searrow 0$ per continuità di h_t (mostrata a (i)).

□

Corollario 1.1. *Sia X una variabile aleatoria Pareto. Allora, per ogni* $1 \leq t \leq 10$, $F_{S(X^n)}(t) - \log t$ *decresce in modo monotono a 0 per* $n \to \infty$.

Dimostrazione. Sia X una Pareto di parametro $\alpha > 0$. Allora $\forall x \geq 1$ segue che

$$\mathbb{P}(X^n > x) = \mathbb{P}(X > x^{\frac{1}{n}}) = x^{-\frac{\alpha}{n}},$$

ossia X^n è una Pareto di parametro $\frac{\alpha}{n}$. Pertanto la conclusione segue dal Teorema 1.3.

□

Quest'ultimo corollario mostra quindi che le funzioni di ripartizione dei significandi delle potenze X_α^n per qualsiasi variabile aleatoria Pareto X_α si avvicinano alla funzione di ripartizione di Benford $\log t$ man mano che $n \to \infty$, e lo fanno in modo monotono. Inoltre, le funzioni di ripartizione di $S(X^n)$ si avvicinano a $\log t$ per ogni variabile casuale X dotata di una densità, come si vedrà nel Teorema 2.7, tuttavia è bene sottolineare che, in generale, questa convergenza non è monotona.

A differenza del Teorema 1.3, nessuna distribuzione uniforme soddisfa alla legge di Benford, indipendentemente da quanto sia ampio il suo range o da dove sia centrata. Questa affermazione può essere quantificata esplicitamente per le distribuzioni uniformi positive come segue. Inoltre, da questo momento in poi, indicheremo con $\langle\cdot\rangle$ la parte frazionaria, ossia

$$\langle\cdot\rangle = \cdot - \lfloor\cdot\rfloor.$$

Teorema 1.4. *Per ogni variabile aleatoria positiva distribuita uniformemente,*

$$\max_{1\le t<10} |F_{S(X_\alpha)}(t) - \log t| \ge \frac{1}{18} + \frac{1}{2}(\log 9 - \log e + \log\log e) = 0.1344,$$

e l'uguaglianza può essere realizzata.

Dimostrazione. Sia X una variabile aleatoria positiva distribuita uniformemente su (0,T), con $T > 0$ (solo tale caso verrà trattato). In prima istanza si noti che $F_{S(X)}(t) = F_{\langle\log X\rangle}(\log t)$ per ogni $1 \le t < 10$, infatti

$$\begin{aligned}
F_{S(X)}(t) &= \mathbb{P}(S(X) \le t) \\
&= \mathbb{P}(10^{\log|X| - \lfloor\log|X|\rfloor} \le t) \\
&\overset{\text{X}\sim\text{ Unif(0,T)}}{=} \mathbb{P}(10^{\log X - \lfloor\log X\rfloor} \le t) \\
&= \mathbb{P}(10^{\langle\log X\rangle} \le t) \\
&= \mathbb{P}(\langle\log X\rangle \le \log t) \\
&= F_{\langle\log X\rangle}(\log t)
\end{aligned}$$

Dunque

$$\max_{1\le t<10} |F_{S(X)}(t) - \log t| = \max_{0\le s\le 1} |F_{\langle\log X\rangle}(s) - s|.$$

Poiché per ipotesi X ha distribuzione uniforme su (0,T) con $T > 0$, la funzione di ripartizione per $\log X$ sarà

$$\begin{aligned}
F_{\log X}(x) &= \mathbb{P}(logX \le x) \\
&= \mathbb{P}(X \le 10^x) \\
&= \begin{cases} \frac{10^x}{T} & \text{se } x < \log T, \\ 1 & \text{se } x \ge \log T, \end{cases}
\end{aligned}$$

e con $\delta := \langle\log T\rangle$ segue che, per ogni $0 \le s \le 1$,

$$\begin{aligned}
F_{\langle\log X\rangle}(s) &= \sum_{k\in\mathbb{Z}} \left(F_{\log X}\,(s+k) - F_{\log X}\,(k)\right) \\
&= \begin{cases} \frac{10}{9}(10^{s-\delta} - 10^{-\delta}) & \text{se } 0 < s \le \delta, \\ 1 - \frac{10}{9}(10^{-\delta} - 10^{s-\delta-1}) & \text{se } \delta \le s \le 1. \end{cases}
\end{aligned}$$

Poniamo ora per semplicità $G_\delta(s) = F_{\langle \log X \rangle}(s) - s$ e si osservi che essa

i. è una funzione continua, perché composizione di funzioni continue;

ii. è differenziabile ovunque eccetto per $s = \delta$, infatti

$$G'_\delta(s) = \begin{cases} \frac{10}{9} 10^{s-\delta} (\log e)^{-1} - 1 & \text{se } 0 < s < \delta, \\ \frac{10}{9} 10^{s-\delta-1} (\log e)^{-1} - 1 & \text{se } \delta < s \leq 1 \end{cases}$$

ma

$$\begin{aligned} \lim_{s \to \delta^-} \frac{10}{9} 10^{s-\delta} (\log e)^{-1} - 1 &= \frac{10}{9} (\log e)^{-1} - 1 \\ &\neq \frac{10}{9} (10 \log e)^{-1} - 1 \\ &= \lim_{s \to \delta^+} \frac{10}{9} 10^{s-\delta-1} (\log e)^{-1} - 1 \end{aligned}$$

e $G_\delta(1) = G_\delta(0) = 0$;

iii. è strettamente convessa negli intervalli $[0, \delta]$ e $[\delta, 1]$, difatti

$$G''_\delta(s) = \begin{cases} \frac{10}{9} 10^{s-\delta} (\log e)^{-2} > 0 & \text{se } 0 < s < \delta, \\ \frac{10}{9} 10^{s-\delta-1} (\log e)^{-2} > 0 & \text{se } \delta < s \leq 1 \end{cases}$$

iv. $G'_\delta(0^+) = \frac{10}{9} \left(10^\delta \log e\right)^{-1} - 1 = G'_\delta(1^-)$, $0 < \delta < 1$.

Questo dimostra che G_δ assume il suo valore massimo in $s = \delta$. Per comodità di notazione, sia

$$\psi(\delta) = G_\delta(s) = \frac{10}{9}(1 - 10^{-\delta}) - \delta$$

e si osservi che ψ è continua, in quanto lo era G_δ, con $\psi(0) = \psi(1) = 0$. Poichè $\psi''(\delta) = -\frac{10}{9} 10^{-\delta} (\log e)^{-1} < 0$ per ogni $0 \leq \delta < 1$, la funzione ψ è strettamente concava e assume massimo in un unico $0 < \delta^* < 1$. Nello specifico, $\delta^* = 1 - \log 9 - \log \log e = 0.4079$. Si assuma ora che $0 \leq \delta \leq \delta^*$. In questo caso la funzione G_δ possiede un minimo non positivo in $s = 1 - \delta - \delta^* > \delta$, con

$$\begin{aligned} G_\delta(1 + \delta - \delta^*) &= 1 - \frac{10}{9}(10^{-\delta^*} - 10^{-\delta}) - 1 - \delta + \delta^* \\ &= \psi(\delta) - \psi(\delta^*); \end{aligned}$$

dunque $\max_{0\leq s\leq 1} G_\delta(s) = \psi(\delta)$ così come $-\min_{0\leq s\leq 1} G_\delta(s) = \psi(\delta^*) - \psi(\delta)$. In modo analogo, se $\delta^* < \delta < 1$ allora G_δ avrà un minimo negativo in $s = \delta - \delta^* < \delta$, con

$$\begin{aligned} G_\delta(\delta - \delta^*) &= \frac{10}{9}(10^{-\delta^*} - 10^{-\delta}) - \delta + \delta^* \\ &= \psi(\delta) - \psi(\delta^*). \end{aligned}$$

Riassumendo quindi, per ogni $0 \leq \delta < 1$,

$$\max_{0\leq s\leq 1} G_\delta(s) = \psi(\delta) \quad e \quad -\min_{0\leq s\leq 1} G_\delta(s) = \psi(\delta^*) - \psi(\delta),$$

e, di conseguenza,

$$\begin{aligned} \max_{0\leq s<1} |F_{\langle \log X\rangle}(s) - s| &= \max_{0\leq s\leq 1} |G_\delta(s)| \\ &= \max\{\psi(\delta), \psi(\delta^*) - \psi(\delta)\} \\ &\geq \frac{1}{2}(\psi(\delta) + \psi(\delta^*) - \psi(\delta)) \\ &= \frac{1}{2}\psi(\delta^*) \\ &= \frac{1}{18} + \frac{1}{2}(1 - \log 9 - \log\log e), \end{aligned}$$

con l'uguaglianza che si verifica esattamente per i due valori di δ tali che $\psi(\delta) = \frac{1}{2}\psi(\delta^*)$. Questo dimostra che l'affermazione vale per un qualsiasi X distribuito uniformemente su (0,T) per un certo $T > 0$.

□

Osservazione 1.2. *Si può vedere facilmente che*

$$max_{1\leq t<10} |F_{S(X)}(t) - \log t| \geq \delta$$

vale per un opportuno valore di $\delta > 0$ ogni volta che la variabile aleatoria X è distribuita uniformemente(su un intervallo aperto che può contenere 0), normalmente o esponenzialmente. Tuttavia in questi casi, a differenza di quanto visto nel Teorema 1.4, il miglior valore di δ, che coincide con il più grande valore di δ che soddisfi ancora la disuguaglianza, non è noto.

Il seguente lemma fornisce un quadro conveniente per studiare le probabilità sulla σ-algebra dei significandi traducendole in misure di probabilità sullo spazio classico dei sottoinsiemi di Borel di [0,1), cioè su $\Big([0,1), \mathcal{B}[0,1)\Big)$. Prima di enunciare tale lemma, ricordiamo che per ogni funzione $f\colon \mathbb{R} \to \mathbb{R}$ con $\sigma(f) \subset \mathcal{A}$ e ogni misura di probabilità $\mathbb{P}$ su $(\Omega, \mathcal{A})$, f e $\mathbb{P}$ insieme inducono una misura di probabilità su $(\mathbb{R}, \mathcal{B})$ in modo naturale, ponendo

$$\mathbb{P}_f(B) = \mathbb{P}(f^{-1}(B)) \quad \forall B \in \mathcal{B}. \tag{1.1}$$

Lemma 1.2. *La funzione $\ell \colon \mathbb{R}^+ \to [0,1)$ definita come $l(x) = \log S(x)$ stabilisce una corrispondenza biunivoca (isomorfismo di misura) tra le misure di probabilità su $(\mathbb{R}^+, \mathcal{S})$ e le misure di probabilità su $\Big([0,1), \mathcal{B}[0,1)\Big)$.*

Dimostrazione. Da $\ell^{-1}([a,b]) = S^{-1}([10^a, 10^b])$ per ogni $0 \leq a < b < 1$ segue che $\sigma(\ell) = \mathbb{R}^+ \cap \sigma(S) = \mathcal{S}$, dunque $\mathbb{P}_l$ definita come in (1.1), con una qualsiasi probabilità $\mathbb{P}$ su $(\mathbb{R}^+, \mathcal{S})$, è una misura di probabilità ben definita su $\Big([0,1), \mathcal{B}[0,1)\Big)$.

D'altra parte, data una qualsiasi misura di probabilità P su $\Big([0,1), \mathcal{B}[0,1)\Big)$ e un A in $\mathcal{S}$, sia $B \in \mathcal{B}[0,1)$ l'unico insieme per cui $A = \bigcup_{k \in \mathbb{Z}} 10^k 10^B$, dove $10^B = \{10^s : s \in B\}$ e definiamo

$$P^\ell(A) = P(B).$$

Si verifica facilmente che $\ell(A) = B$, $\ell^{-1}(B) = A$ e che P^ℓ è una misura di probabilità ben definita su $(\mathbb{R}^+, \mathcal{S})$. Inoltre,

$$(P^\ell)_\ell(B) = P^\ell(\ell^{-1}(B)) = P^\ell(A) = P(B) \;\; \forall B \in \mathcal{B}[0,1),$$

ossia $(P^\ell)_\ell = \mathbb{P}$, dunque ogni misura di probabilità su $\Big([0,1), \mathcal{B}[0,1)\Big)$ è della forma $\mathbb{P}_\ell$ per un'opportuna $\mathbb{P}$. D'altro canto,

$$(\mathbb{P}_\ell)^\ell(A) = \mathbb{P}_\ell(B) = \mathbb{P}(\ell^{-1}(B)) = \mathbb{P}(A) \text{ per ogni } A \in \mathcal{S};$$

pertanto, $(\mathbb{P}_\ell)^\ell = \mathbb{P}$, e la corrispondenza $\mathbb{P} \mapsto \mathbb{P}_\ell$ è anch'essa iniettiva. Dunque, riassumendo, la mappa $\mathbb{P} \leftrightarrow \mathbb{P}_\ell$ è biettiva.

$\square$

A questo punto sorge spontanea una domanda: nella dimostrazione del Lemma 1.2 non avremmo potuto costruire la corrispondenza tra le misure di probabilità su $(\mathbb{R}^+, \mathcal{S})$ e quelle su $\Big([0,1), \mathcal{B}[0,1)\Big)$ in un altro modo? Appare subito chiaro che, per esempio, avremmo potuto usare la funzione $\tilde{\ell}(x) = \frac{1}{9}(S(x) - 1)$ invece di ℓ. Il perché della scelta di una tale ℓ diventa però evidente una volta che si pone l'attenzione sulla sua relazione con la legge di Benford. Ricordiamo infatti la Definizone 1.4: "$\mathbb{B}$ è l'unica misura di probabilità su $(\mathbb{R}^+, \mathcal{S})$ tale che

$$\mathbb{B}(\{x > 0 : S(x) \leq t\}) = \mathbb{B}\left(\bigcup_{k \in \mathbb{Z}} 10^k [1, t]\right) = \log t \quad \forall t \in [1, 10)."$$

Fino a questo momento si è visto come la misura di probabilità $\mathbb{B}$ su $(\mathbb{R}^+, \mathcal{S})$ è la formalizzazione più naturale della legge di Benford, tuttavia su $\Big([0,1), \mathcal{B}[0,1)\Big)$

la distribuzione uniforme $\lambda_{0,1}$ ha molte proprietà speciali che le conferiscono un ruolo molto importante. La rilevanza della scelta specifica per ℓ nel Lemma 1.2 si spiega pertanto con il fatto che $\mathbb{B}_\ell = \lambda_{0,1}$.

1.1 Distribuzione Uniforme e Legge di Benford

In questo paragrafo si affronteranno i principali risultati riguardanti la caratterizzazione della legge di Benford tramite la distribuzione uniforme, i quali verranno spesso richiamati nel Capitolo 2.

Definizione 1.6. *Una sequenza (x_n) di numeri reali è uniformemente distribuita modulo 1 se*

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \langle x_n \rangle \le s\}}{N} = s \ \forall s \in [0,1);$$

una variabile aleatoria X sullo spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ è u.d. mod 1 se

$$\mathbb{P}\left(\langle X \rangle \le s\right) = s \ \forall s \in [0,1);$$

una misura di probabilità P su $(\mathbb{R}, \mathcal{B})$ è u.d. mod 1 se

$$P\left(\{x : \langle x \rangle \le s\}\right) = P\left(\bigcup_{k\in\mathbb{Z}} [k, k+s]\right) = s \ \forall s \in [0,1).$$

Osservazione 1.3. *Equivalentemente possiamo dire che una sequenza (x_n) di numeri reali è uniformemente distribuita modulo 1 se*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\langle x_n \rangle \le s}(n) = s \quad \forall t \in [0,1).$$

Il prossimo risultato è un risultato centrale nella teoria della legge di Benford e la sua importanza risiede nel fatto che permette di applicare la potente teoria della distribuzione uniforme modulo uno.
Si ricordino le convenzioni $\log 0 = 0$ e $S(0) = 0$.

Teorema 1.5. *Una variabile aleatoria è Benford se e solo se la variabile aleatoria $\log|X|$ è uniformemente distribuita modulo 1.*
Una successione di numeri reali (o una misura di probabilità di Borel) è Benford se e solo se il logaritmo decimale del suo valore assoluto è u.d. mod 1.

Dimostrazione. Sia X una variabile aleatoria. Allora, per ogni $s \in [0,1)$

$$\begin{aligned}
\mathbb{P}(\langle \log|X| \rangle \leq s) &= \mathbb{P}\left(\log|X| \in \bigcup_{k\in\mathbb{Z}} [k, k+s] \right) \\
&= \mathbb{P}\left(|X| \in \bigcup_{k\in\mathbb{Z}} [10^k, 10^{k+s}] \right) + \mathbb{P}(X=0) \\
&= \mathbb{P}\left(S(X) \leq 10^s \right).
\end{aligned}$$

Dalle Definizioni 1.3 e 1.5 segue quindi che X è Benford se e solo se $\mathbb{P}(S(X) \leq 10^s) = \log 10^s = s$ per ogni $s \in [0,1)$, cioè X è Benford se e solo se $\mathbb{P}(\langle \log|X| \rangle \leq s) = s$, ossia $\log|X|$ è u.d. mod 1.

□

La seguente proposizione che proviene dalla teoria di base della distribuzione uniforme modulo 1, insieme al Teorema 1.5, sarà fondamentale per stabilire la proprietà Benford per molte successioni e variabili aleatorie.

Proposizione 1.2. *i. La successione (x_n) è u.d. mod 1 se e solo se $(kx_n + b)$ è u.d. mod 1 per ogni intero $k \neq 0$ ed ogni $b \in \mathbb{R}$. Inoltre, (x_n) è u.d. mod 1 se e solo se (y_n) è u.d. mod 1 ogni qualvolta $\lim_{n\to\infty}|y_n - x_n| = 0$.*

ii. Una variabile aleatoria X è u.d. mod 1 se e solo se $kX + b$ è u.d. mod 1 per ogni intero $k \neq 0$ e $b \in \mathbb{R}$.

Vediamo ora due risultati che ci torneranno utili per dimostrare il Teorema 1.7, il quale fornisce una condizione sufficiente e necessaria affinchè una successione avente comportamento asintotico esponenziale sia Benford.

Teorema 1.6. *Si considerino (a_n) e (b_n) due successioni di numeri reali con $|a_n| \to \infty$ per $n \to \infty$ e tali che $\lim_{n\to\infty}|a_n/b_n|$ esiste ed è positivo. Allora (a_n) è Benford se e solo se (b_n) è Benford.*

Dimostrazione. Possiamo supporre, senza perdità di generalità, che $\lim_{n\to\infty}|a_n/b_n| = 1$. Di conseguenza, poiché per ipotesi $|a_n| \to \infty$ per $n \to \infty$, si avrà che $|b_n| \to \infty$ per $n \to \infty$ e $|\log|a_n| - \log|b_n|| = |\log|a_n/b_n|| \to 0$ per $n \to \infty$ (per ipotesi $\lim_{n\to\infty}|a_n/b_n|$). Se (a_n) è Benford allora, per il Teorema 1.5, $(\log|a_n|)$ è u.d mod 1. Dalla Proposizione 1.2 (i) segue quindi che anche $(\log|a_n|)$ è u.d mod 1, di conseguenza, per il Teorema 1.5, (b_n) è Benford. Poiché i ruoli di (a_n) e (b_n) sono interscambiabili, ho dimostrato quanto voluto.

□

La prossima proposizione raccoglie diversi risultati di base provenienti dalla teoria della distribuzione uniforme per successioni.

Proposizione 1.3. *Si consideri una successione di numeri reali* $(x_n) = (x_1, x_2, \dots)$.

i. *Se* (x_n) *è periodica, ovvero* $x_{n+p} = x_n$ *per un certo* $p \in \mathbb{N}$*, allora* $(n\theta + x_n)$ *è u.d. mod 1 se e solo se* θ *è irrazionale.*

ii. *La successione* (x_n) *è u.d. mod 1 se e solo se* $(x_n + a \log n)$ *è u.d. mod 1 per ogni* $a \in \mathbb{R}$.

iii. *Se* $\lim_{n\to\infty}(x_{n+1} - x_n) = \theta$ *per un qualche* θ *irrazionale, allora* (x_n) *è u.d. mod 1.*

Osservazione 1.4. *Nella precedente proposizione non è possibile sostituire la successione* $(n\theta)$ *con una qualunque successione uniformemente distribuita* (θ_n)*. In altre parole,* $(\theta_n + x_n)$ *potrebbe non essere uniformemente distribuita modulo 1 sebbene* (θ_n) *sia u.d. mod 1 e* (x_n) *sia una successione periodica.*

Teorema 1.7. *Sia* (b_n) *una successione di numeri reali tale che* $\lim_{n\to\infty}|b_n/a^n|$ *esiste ed è positivo per un certo* $a > 0$*. Allora* (b_n) *è Benford se e solo se* $\log|a|$ *è irrazionale.*

Dimostrazione. (b_n) è Benford $\overset{\text{Teorema 1.6}}{\iff}$ (a^n) è Benford

$$\overset{\text{Teorema 1.5}}{\iff} (n \log a) \text{ è u.d mod 1} \overset{\text{Proposizione 1.3(i)}}{\iff} \log a \text{ è irrazionale.}$$

□

Prima di enunciare il Teorema 1.8, introduciamo un lemma che raccoglie molte delle più importanti proprietà dei coefficienti di Fourier: tali risultati verranno sfruttati numerose volte anche nel Capitolo 2.

Si ricordi che per ogni variabile aleatoria X con valori in $[0, 1)$, o, in modo equivalente, per la misura di probabilità associata P_X su $([0, 1), \mathcal{B}[0, 1))$, i coefficienti di Fourier (o coefficienti di Fourier–Stieltjes) di X e P_X sono la successione di numeri complessi limitata e che si estende all'infinito in entrambe le direzioni $\left(\widehat{P_X}(k)\right)_{k\in\mathbb{Z}}$, data da

$$\begin{aligned}
\widehat{P_X}(k) &= \mathbb{E}[e^{2\pi i k X}] \\
&= \int_0^1 e^{2\pi i k s}\, dP_X(s) \\
&= \int_0^1 \cos(2\pi k s)\; dP_X(s) + i \int_0^1 \sin(2\pi k s)\; dP_X(s), \;\; k \in \mathbb{Z}.
\end{aligned}$$

Lemma 1.3. *Si considerino X e X_n, $n \in \mathbb{N}$, delle variabili aleatorie a valori in $[0,1)$, aventi distribuzioni P_X e P_{X_n} e coefficienti di Fourier $\left(\widehat{P_X}(k)\right)_{k\in\mathbb{Z}}$ e $\left(\widehat{P_{X_n}}(k)\right)_{k\in\mathbb{Z}}$ rispettivamente. Allora*

i $|\widehat{P_X}(k)| \leq 1$ *per ogni* $k \in \mathbb{Z}$ *e* $\widehat{P_X}(0) = 1$;

ii Se $|\widehat{P_X}(k)| = 1$ *per qualche* $k \neq 0$ *allora* $\mathbb{P}(kX \in s + \mathbb{Z}) = 1$ *per qualche* $s \in [0,1)$;

iii $(\widehat{P_X}(k))_{k\in\mathbb{Z}}$ *determina univocamente* P_X, *ossia* $\widehat{P_{X_1}}(k) = \widehat{P_{X_2}}(k)$ *per ogni* $k \in \mathbb{Z}$ *implica che* $P_{X_1} = P_{X_2}$;

iv Se X_1 *e* X_2 *sono indipendenti allora* $\widehat{P_{\langle X_1+X_2\rangle}}(k) = \widehat{P_{X_1}}(k) \cdot \widehat{P_{X_2}}(k)$ *per ogni* $k \in \mathbb{Z}$;

v $\lim_{n\to\infty} \widehat{P_{X_n}}(k) = \widehat{P_X}(k)$ *per ogni* $k \in \mathbb{Z}$ *se e solo se*

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \leq x\right) = \mathbb{P}\left(X \leq x\right) \text{ per ogni } x \in \mathbb{R} \text{ con } \mathbb{P}\left(X = x\right) = 0.$$

vi X *è uniforme, ossia* $X = U(0,1)$ *se e solo se*

$$\widehat{P_{\langle X\rangle}}(k) = \widehat{\lambda_{0,1}}(k) = \begin{cases} 1 & \text{se } k = 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Teorema 1.8. *Siano X, Y variabili aleatorie. Allora:*

i. Se X è u.d. mod 1 e Y è indipendente da X si ha che $X + Y$ è u.d. mod 1;

ii. Se $\langle X\rangle$ e $\langle X+a\rangle$ hanno la stessa distribuzione per un qualche $a \in \mathbb{R}\setminus\mathbb{Q}$ allora X è u.d. mod 1.

Dimostrazione. (i) Supponiamo che X sia u.d. mod 1 e che Y sia indipendente. Per il Lemma 1.3 (vi), $\widehat{P_X}(k) = 0$ per ogni $k \neq 0$, dunque per il Lemma 1.3 (iv) si ha che

$$\widehat{P_{\langle X+Y\rangle}}(k) = \widehat{P_X}(k) \cdot \widehat{P_Y}(k) = 0 \text{ per ogni } k \neq 0$$

e $\widehat{P_{\langle X+Y\rangle}}(0) = 1$ (Lemma 1.3 (i)). Pertanto, per il Lemma 1.3 (iii,iv) segue che $\langle X + Y\rangle = U(0,1)$, cioè $X + Y$ è u.d. mod 1.
(ii) Supponiamo che $\langle X\rangle$ e $\langle X + a\rangle$ abbiano la stessa distribuzione per un

qualche $a \in \mathbb{R} \setminus \mathbb{Q}$. Osserviamo che se $Y = a$ con probabilità 1 allora $\widehat{P_{\langle Y \rangle}}(k) = e^{2\pi i k a}$ per ogni $k \in \mathbb{Z}$. Di conseguenza, se $\langle X \rangle$ e $\langle X + a \rangle$ hanno la stessa distribuzione, allora

$$\widehat{P_{\langle X \rangle}}(k) = \widehat{P_{\langle X+a \rangle}}(k) \overset{\text{Lemma 1.3 (iv)}}{=} \widehat{P_{\langle X \rangle}}(k) \cdot e^{2\pi i k a},$$

ma se a è irrazionale avrò che $e^{2\pi i k a} \neq 1$ per ogni $k \neq 0$ e dunque, affinché sia verificata l'identità, si deve avere che $\widehat{P_{\langle X \rangle}}(k) = 0$ per ogni $k \neq 0$. Per il Lemma 1.3 (iii,vi) dunque, $\widehat{P_{\langle X \rangle}} = \widehat{\lambda_{0,1}}$, cioè $X = U(0, 1)$.

$\square$

Capitolo 2

Processi stocastici a valori reali

2.1 Convergenza delle variabili aleatorie alla legge di Benford

L'analisi delle successioni di variabili aleatorie, con particolare attenzione al caso specifico di prodotti e somme di variabili aleatorie indipendenti e identicamente distribuite (i.i.d.), rappresenta un tema centrale nella teoria classica delle probabilità. Questa sezione studierà scenari generali in cui la legge di Benford emerge come una distribuzione "attrattiva", nel senso che anche quando i dati iniziali non seguono questa legge, certe operazioni su di essi (ad esempio combinazioni lineari o trasformazioni non lineari) li portano a conformarsi alla legge di Benford. L'obiettivo è inoltre individuare alcune delle principali proprietà limite Benford relative alle successioni di variabili aleatorie. Queste proprietà includono i tre fatti fondamentali seguenti:

i. le potenze di ogni variabile aleatoria continua convergono alla legge di Benford;

ii. i prodotti di campioni casuali estratti da qualsiasi distribuzione continua convergono alla legge di Benford;

iii. se i campioni casuali sono presi da distribuzioni casuali scelte senza condizionamenti, allora il campione combinato ottenuto converge alla legge di Benford.

Per definizione, un campione casuale è una sequenza finita $X_1, X_2, \dots, X_n$ di variabili aleatorie i.i.d.
Si ricordi che una successione di variabili aleatorie $(X_n) = (X_1, X_2, \dots)$ converge in distribuzione alla variabile aleatoria X $\left(X_n \xrightarrow{\mathcal{D}} X\right)$ se

$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$ per ogni $x \in \mathbb{R}$ per cui $\mathbb{P}(X = x) = 0$. Un altro concetto importante nei teoremi limite della teoria della probabilità è la convergenza quasi certa: una successione (X_n) converge a X quasi certamente (q.c.) se $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$ e scriveremo che $X_n \overset{q.c.}{\to} X$. Si può facilmente dimostrare che la convergenza quasi certa implica la convergenza in distribuzione, ma, in generale, non vale l'implicazione inversa. Enunciamo ora 3 teoremi cardine della teoria della probabilità che saranno fondamentali in alcune dimostrazioni di questa sezione.

Teorema 2.1 (Teorema della Convergenza Dominata). *Se $X_n \xrightarrow{n\to\infty} X$ in probabilità ed esiste $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ tale che $|X_n| \leq Y$ P-quasi certamente per ogni $n \in \mathbb{N}$ allora*

$$\mathbb{E}[X_n] \xrightarrow{n\to\infty} \mathbb{E}[X].$$

Teorema 2.2 (Legge Forte dei Grandi Numeri). *Siano $X_1, X_2, \dots$ variabili aleatorie reali in $L^2(\Omega, \mathcal{A}, \mathbb{P})$ identicamente distribuite e a due a due indipendenti. Allora*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} \mathbb{E}[X_1] \quad \mathbb{P}\text{-}\textit{quasi certamente.}$$

Teorema 2.3 (Teorema del Limite Centrale). *Sia $(X_i)_{i\in\mathbb{N}} \subset L^2(\Omega, \mathcal{A}, \mathbb{P})$ una successione di variabili aleatorie i.i.d. Poniamo $m = \mathbb{E}[X_1]$, $\sigma^2 = Var(X_1)$ e $Z_n = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}(X_i - m)\right)$. Allora si ha che $(Z_n)_{n\in\mathbb{N}}$ converge in distribuzione ad una variabile aleatoria con legge gaussiana con media 0 e varianza σ^2.*

La prossima definizione adatta le nozioni generali di convergenza in distribuzione e convergenza quasi certa alla legge di Benford.

Definizione 2.1. *Sia $(X_n) = (X_1, X_2, \dots)$ una sequenza di variabili aleatorie a valori reali.*

i. La sequenza (X_n) converge in distribuzione alla legge di Benford se

$$lim_{n\to\infty}\mathbb{P}(S(X_n) \leq t) = \log t \textit{ per ogni } t \in [1, 10),$$

o, equivalentemente, se $S(X_n) \overset{D}{\to} S(X)$, dove X è una variabile aleatoria Benford.

ii. La sequenza (X_n) è Benford con probabilità uno se

$$\mathbb{P}((X_n) \textit{ è una sequenza Benford}) = 1.$$

Si noti che la funzione di ripartizione della legge di Benford, $F_{\mathbb{B}}(t) = \log t$, è continua. Di conseguenza, se (X_n) converge in distribuzione alla legge di Benford allora $|F_{S(X_n)}(t) - \log t| \to 0$ uniformemente su [1,10).
Osserviamo inoltre che le due nozioni di conformità alla legge di Benford presentate nella Definizione 2.1 sono in generale non correlate, vediamolo con due esempi.
Esempio 2.1
Sia X una variabile aleatoria Benford e, per ogni $n \in \mathbb{N}$, definiamo X_n come segue:

$$X_n = \min\{S(X), 10 - n^{-1}\}.$$

Innanzitutto si noti che che $S(X_n) = X_n$ e dunque per ogni $t \in [1, 10)$ e $n \in \mathbb{N}$,

$$\begin{aligned} |F_{S(X_n)}(t) - \log t| &= |\mathbb{P}(S(X_n) \le t) - \log t| \\ &\overset{S(X_n)=X_n}{=} |\mathbb{P}(X_n \le t) - \log t| \end{aligned}$$

Prima di continuare osserviamo che

$$\begin{aligned} \mathbb{P}(X_n \le t) &= 1 - \mathbb{P}(X_n > t) \\ &= 1 - \mathbb{P}\left(\min\{S(X), 10 - n^{-1}\} > t\right) \\ &= 1 - \mathbb{P}\left(S(X) > t, 10 - n^{-1} > t\right) \\ &\overset{indipendenza}{=} 1 - (1 - \mathbb{P}\left(S(X) \le t\right))\, \mathbb{P}\left(10 - n^{-1} > t\right). \end{aligned}$$

Di conseguenza

- se $1 \le t \le 10 - n^{-1}$: $\mathbb{P}(X_n \le t) = 1 - (1 - \log t) \cdot 1 = \log t$;
- se $10 - n^{-1} \le t \le 10$: $\mathbb{P}(X_n \le t) = 1 - \mathbb{P}(S(X_n) > t) \cdot 0 = 1$.

Pertanto

$$\begin{aligned} |F_{S(X_n)}(t) - \log t| &= \begin{cases} 0 \text{ se } 1 \le t < 10 - n^{-1}, \\ 1 - logt \text{ se } 10 - n^{-1} \le t < 10, \end{cases} \\ &\le 1 - \log(10 - n^{-1}) \\ &< \frac{1}{20} n^{-1}, \end{aligned}$$

dunque $|F_{S(X_n)} - \log t| \to 0$ uniformemente su [1,10], ovvero (X_n) converge in distribuzione alla legge di Benford. D'altra parte, $X_n = S(X)$ ogni qual volta $S(X) \le 10 - n^{-1}$, ma poiché $10 - n^{-1} \to 10$ per $n \to \infty$ e S(X) assume valori tra 1 e 9, esisterà un certo $n_0 \in \mathbb{N}$ tale che

$S(X) \leq 10 - n^{-1}$ $\forall n \geq n_0$, ovvero definitivamente $X_n = S(X)$. Pertanto la sequenza (X_n) è costante con probabilità 1 e, dunque, non è Benford. In altre parole, $\mathbb{P}\left((X_n)\right.$ è una sequenza Benford$) = 0$.

Esempio 2.2 Sia X una variabile aleatoria tale che $\mathbb{P}(X = 2) = 1$ e per ogni $n \in \mathbb{N}$ definiamo $X_n = X^n$, quindi $X_n = 2^n$. Per il Teorema 1.8, essendo $\log 2$ irrazionale, la sequenza (X_n) è Benford con probabilità 1 ma, per ogni $n \in \mathbb{N}$, la distribuzione di $S(X_n)$ è concentrata nel punto $S(2^n) = 10^{\langle nlog2 \rangle}$, cioè $\mathbb{P}\left(S(X_n) = 10^{\langle nlog2 \rangle}\right) = 1$, perciò

$$F_{S(X_n)}(t) = \begin{cases} 0 & \text{se } 1 \leq t < 10^{\langle n \log 2 \rangle} \\ 1 & \text{se } 10^{\langle n \log 2 \rangle} \leq t < 10, \end{cases}$$

da cui segue che

$$\sup_{t \in [1,10)} |F_{S(X_n)}(t) - \log t| \geq \frac{1}{2},$$

dunque (X_n) non converge in distribuzione alla legge di Benford.

A questo punto è utile riformulare alcuni risultati del capitolo precedente utilizzando la terminologia della Definizione 2.1. In questo contesto, il Corollario 1.1 afferma semplicemente che se X è una variabile aleatoria Pareto, allora la sequenza di potenze (X^n) converge in distribuzione alla legge di Benford per $n \to \infty$.

Un ulteriore esempio di convergenza in distribuzione alla legge di Benford, che si ottiene facilmente dai risultati del Capitolo 1, è il seguente teorema. Prima di enunciarlo, si ricordi che la mediana di una variabile aleatoria X (med$X$) è un qualsiasi numero $m \in \mathbb{R}$ tale che $\mathbb{P}(X \leq m) \geq \frac{1}{2}$ e $\mathbb{P}(X \geq m) \geq \frac{1}{2}$.

Teorema 2.4. *Sia (X_n) una sequenza di variabili aleatorie Pareto. Allora (X_n) converge in distribuzione alla legge di Benford se e solo se med$X_n \to \infty$ per $n \to \infty$.*

Dimostrazione. Per ogni $n \in \mathbb{N}$, sia X_n una Pareto di parametro $\alpha_n > 0$. Si indichi con $m_n = \text{med}X_n$ la mediana di X_n (la quale è unica). Allora, per definizione di mediana e di distribuzione Pareto, si ha che
$\mathbb{P}(X_n > m_n) = \frac{1}{2} = m_n^{-\alpha_n}$, pertanto $\alpha_n = \log 2/ \log m_n$. Dunque, $\alpha_n \to 0$ se e solo se $m_n \to \infty$ e la conclusione segue dal Teorema 1.3, dove la convergenza monotona di $F_{S(X_\alpha)}(t)$ a $\log t$ per $\alpha \to 0$ viene utilizzata per dimostrare il "solo se".

$\square$

2.2 Potenze, prodotti e somme di variabili aleatorie

Il Teorema 1.7 ci fornisce l'esempio più semplice di una successione di Benford, ovvero la successione $(a^n) = (a, a^2, a^3, \dots)$, la quale è Benford se e solo se $\log|a|$ è irrazionale. Questo esempio può essere facilmente interpretato in termini probabilistici: la successione (deterministica) (a^n) coincide con probabilità uno con la successione di variabili aleatorie $(X_1, X_1X_2, X_1X_2X_3 \dots)$, dove le X_j hanno tutte la stessa distribuzione, ovvero $\mathbb{P}(X_j = a) = 1$. Pertanto, una naturale generalizzazione di (a_n) in termini aleatori è fornita dalla classe di successioni di variabili aleatorie

$$\Big(\prod_{j=1}^{n} X_j\Big) = (X_1, X_1X_2, \dots) \text{ con } X_1, X_2, \dots \text{ identicamente distribuite.} \tag{2.1}$$

Le successioni casuali nella forma (2.1) possono manifestare una grande varietà di comportamenti. In questa sezione vengono analizzate le proprietà di Benford di (2.1) in due importanti casi particolari:

i. i fattori X_j sono identici (quindi dipendenti), cioè con probabilità uno si ha che $X_j = X_1$ per ogni j;

ii. i fattori X_j sono indipendenti, ovvero $X_1, X_2, \dots$ sono variabili aleatorie indipendenti ed identicamente distribuite.

Si tenga a mente che la successione deterministica (a^n) è un caso speciale appartenente ad entrambi gli scenari.

Potenze di una variabile aleatoria

Se tutti i fattori $X_1, X_2, \dots$ sono identici, allora il processo stocastico descritto in (2.1) assume semplicemente la forma (X^n), dove X è una qualsiasi variabile aleatoria a valori reali. Per tali sequenze, è semplice caratterizzare la conformità alla legge di Benford nel senso della Definizione 2.1.

Osservazione 2.1. *$e^{2\pi ik} = \cos(2\pi ik) + i\sin(2\pi ik) = 1\ \forall k \in \mathbb{Z}$, da cui segue che*

$$e^{2\pi i\langle x\rangle} = e^{2\pi ix}e^{-2\pi i\lfloor x\rfloor} \overset{\lfloor x\rfloor\in\mathbb{Z}}{=} e^{2\pi ix},$$

ovvero la parte intera inferiore di x viene "ignorata" dalla funzione esponenziale complessa.

Teorema 2.5. *Sia X una variabile aleatoria. Allora (X^n) converge in distribuzione alla legge di Benford se e solo se $\lim_{n\to\infty} \mathbb{E}[e^{2\pi in\log|X|}] = 0$.*

Dimostrazione. Definiamo (Y_n) come $Y_n = \log(S(X^n)) = \log\left(10^{\langle n \log|X|\rangle}\right) = \langle n \log|X|\rangle$. Mostriamo che (X^n) converge in distribuzione alla legge di Benford se e solo se $Y_n \xrightarrow{\mathcal{D}} U(0,1)$.

$\Rightarrow$ Supponiamo che (X^n) converga in distribuzione alla legge di Benford, ossia

$$\lim_{n\to\infty} \mathbb{P}(S(X^n) \leq t) = \log t \ \forall t \in [1,10)$$

vogliamo dimostrare che (Y_n) converge in distribuzione a $U(0,1)$, ovvero che

$$\lim_{n\to\infty} \mathbb{P}(Y_n \leq x) = \mathbb{P}(U(0,1) \leq x) \ \forall x \in \mathbb{R} \text{ tale che } \mathbb{P}\left(U(0,1) = x\right) = 0.$$

Partiamo con il notare che $\mathbb{P}\left(U(0,1) = x\right) = 0 \ \forall x \in \mathbb{R}$ e che

$$\mathbb{P}(U(0,1) \leq x) = \begin{cases} 0 \text{ se } x < 0, \\ x \text{ se } 0 \leq x < 1, \\ 1 \text{ se } x \geq 1. \end{cases}$$

Si ha che

- per $x \in [0,1)$

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(Y_n \leq x) &= \lim_{n\to\infty} \mathbb{P}(\log\left(S(X^n)\right) \leq x) \\ &= \lim_{n\to\infty} \mathbb{P}(S(X^n) \leq 10^x) \\ &\overset{10^x \in [1,10)}{=} \log(10^x) \\ &= x \\ &= \mathbb{P}(U(0,1) \leq x);\end{aligned}$$

- per $x < 0$

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(Y_n \leq x) &= \lim_{n\to\infty} \mathbb{P}(S(X^n) \leq 10^x) \\ &\overset{x \leq 0}{=} \lim_{n\to\infty} \mathbb{P}(S(X^n) < 1) \\ &= 0 \\ &= \mathbb{P}(U(0,1) \leq x);\end{aligned}$$

- per $x \geq 1$

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(Y_n \leq x) &= \lim_{n\to\infty} \mathbb{P}(S(X^n) \leq 10^x) \\ &\overset{x \leq 0}{=} \lim_{n\to\infty} \mathbb{P}(S(X^n) \leq 10) \\ &= 1 \\ &= \mathbb{P}(U(0,1) \leq x),\end{aligned}$$

quindi (Y_n) converge in distribuzione a $U(0,1)$.
$\Leftarrow$ Supponiamo ora che (Y_n) converga in distribuzione a $U(0,1)$ e mostriamo che allora (X^n) converge in distribuzione alla legge di Benford. Si ha che, per $t \in [1,10)$,

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(S(X^n) \le t) &= \lim_{n\to\infty} \mathbb{P}(\log(S(X^n)) \le \log t) \\
&= \lim_{n\to\infty} \mathbb{P}(Y_n \le \log t) \\
&\overset{Y_n \xrightarrow{\mathcal{D}} U(0,1)}{=} \mathbb{P}(U(0,1) \le \log t) \\
&\overset{\log t \in [0,1)}{=} \log t,
\end{aligned}$$

cioè (X^n) converge in distribuzione alla legge di Benford.
Dunque ci basta mostrare che $Y_n \xrightarrow{\mathcal{D}} U(0,1)$ se e solo se $\lim_{n\to\infty} \mathbb{E}[e^{2\pi i n \log|X|}] = 0$.
Per il Lemma 1.3(v), $Y_n \xrightarrow{\mathcal{D}} U(0,1)$ è equivalente a dire che

$$\lim_{n\to\infty} \widehat{P_{Y_n}}(k) = \widehat{P_{U(0,1)}}(k).$$

Inoltre, per ogni $k \in \mathbb{Z} \setminus \{0\}$

$$\begin{aligned}
\widehat{P_{U(0,1)}}(k) &= \int_0^1 e^{-2\pi i k x}\, dx \\
&= \frac{e^{-2\pi i k} - e^0}{-2\pi i k} \\
&= 0,
\end{aligned}$$

dunque $Y_n \xrightarrow{\mathcal{D}} U(0,1)$ è equivalente a dire che $\lim_{n\to\infty} \widehat{P_{Y_n}}(k) = 0$.
Supponiamo quindi che $\lim_{n\to\infty} \mathbb{E}[e^{2\pi i n \log|X|}] = 0$ allora, per ogni $k > 0$,

$$\widehat{P_{Y_n}}(k) = \mathbb{E}[e^{2\pi i k Y_n}] \overset{\text{Oss 2.1}}{=} \mathbb{E}[e^{2\pi i k n \log|X|}] \overset{n\to\infty}{\to} 0.$$

Essendo (Y_n) a valori reali $\widehat{P_{Y_n}}(-k) = \overline{\widehat{P_{Y_n}}(-k)}$, pertanto $Y_n \xrightarrow{\mathcal{D}} U(0,1)$.
A questo punto dimostraziamo l'implicazione inversa e supponiamo che $\limsup_{n\to\infty} \left|\mathbb{E}\left[e^{2\pi i n \log|X|}\right]\right| > 0$, ma allora esisterà un $\delta > 0$ e una successione crescente (n_j) di interi positivi tale che $|\mathbb{E}[e^{2\pi i n_j \log|X|}]| \ge \delta$ per ogni j. Da ciò segue che, per ogni $j \in \mathbb{N}$,

$$|\widehat{P_{Y_{n_j}}}(1)| = \left|\mathbb{E}[e^{2\pi i Y_{n_j}}]\right| = \left|\mathbb{E}[e^{2\pi i n_j \log|X|}]\right| \ge \delta > 0,$$

cioè $\widehat{P_{Y_n}}(k)$ non tende a 0 per $n \to \infty$ per $k = 1$, ovvero (Y_n) non converge in distribuzione a $U(0,1)$, pertanto, per quanto visto in precedenza, (X^n) non converge in distribuzione alla legge di Benford.

□

Teorema 2.6. *Sia X una variabile aleatoria. Le seguenti affermazioni sono equivalenti:*

i. *La succesione* (X^n) *è Benford con probabilità uno;*

ii. $\mathbb{P}(\log|X| \in \mathbb{Q}) = 0$, *ovvero* $\log|X|$ *è irrazionale con probabilità uno;*

iii. $\mathbb{P}(S(X^m) = 1) = 0 \ \forall m \in \mathbb{N}$.

Dimostrazione. Si denoti con $(\Omega, \mathcal{A}, \mathbb{P})$ lo spazio di probabilità in cui è definita X. Dal Teorema 1.7, per ogni $w \in \Omega$, la successione

$$(X(w)^n) = (X(w), X(w)^2, X(w)^3, \dots)$$

è Benford se e solo se $\log|X(w)|$ è irrazionale, il che mostra (i) $\iff$ (ii). Per vedere l'equivalenza tra (ii) e (iii) basta osservare semplicemente che per ogni $m \in \mathbb{N}$ e $t \in [1, 10)$

$$\{w \in \Omega : S(X(w)^m) = t\} = \left\{w \in \Omega : \log|X(w)| \in \frac{\log t}{m} + \frac{1}{m}\mathbb{Z}\right\}, \tag{2.2}$$

e quindi $\sup_{m \in \mathbb{N}} \mathbb{P}(S(X^m) = 1) = \mathbb{P}(\log|X| \in \mathbb{Q})$.

□

L'Esempio 2.2 mostra che una sequenza (X^n) può essere Benford con probabilità uno, ma non convergere in distribuzione alla legge di Benford. Tuttavia, l'implicazione inversa è sempre verificata.

Corollario 2.1. *Sia X una variabile aleatoria. Se la sequenza* (X^n) *converge in distribuzione alla legge di Benford allora è Benford con probabilità uno.*

Dimostrazione. Supponiamo che $\mathbb{P}((X_n)$ è una sequenza Benford$) < 1$. Per il Teorema 2.5 si avrebbe allora che $\mathbb{P}(\log|X| \in \mathbb{Q}) > 0$ quindi esisterebbero un $p \in \mathbb{Z}$ e un $q \in \mathbb{Z}$ tali che $\mathbb{P}(|X| = 10^{p/q}) > 0$, ma allora

$$\begin{aligned} F_{S(X^{nq})}(1) &= \mathbb{P}(S(X^{nq}) \leq 1) \\ &\geq \mathbb{P}(|X| = 10^{p/q}) \\ &> 0 \text{ per ogni } n \in \mathbb{N}, \end{aligned}$$

il che mostra che $F_{S(X^n)}(1)$ non converge a $0 = \log 1$, pertanto (X^n) non converge in distribuzione alla legge di Benford.

□

Prima di enunciare il prossimo teorema, ricordiamo un altro risultato fondamentale per quanto riguarda i coefficienti di Fourier.

Lemma 2.1 (Riemann-Lebesgue). *Sia $f \in L^1(\mathbb{T})$, ossia f è una funzione integrabile su $\mathbb{T}$. Allora*

$$\lim_{n\to\infty} \widehat{f}(n) = 0,$$

dove con $\widehat{f}(n)$ denotiamo i coefficienti di Fourier di f.

Teorema 2.7. *Se la variabile aleatoria X è dotata di una densità allora la successione (X^n) converge in distribuzione alla legge di Benford ed è Benford con probabilità uno.*

Dimostrazione. Se X ha una densità, allora ciò varrà anche per $Y = \langle \log|X| \rangle$. Y è integrabile, pertanto possiamo applicare il Lemma di Riemann-Lebesgue, il quale afferma che

$$\lim_{n\to\infty} \widehat{P_Y}(n) = \lim_{n\to\infty} \mathbb{E}[e^{2\pi i n Y}] = 0.$$

Per come è stata definita Y si ha che $\mathbb{E}[e^{2\pi i n Y}] \overset{\text{Oss 2.1}}{=} \mathbb{E}[e^{2\pi i n \log|X|}]$, pertanto, per il Teorema 2.5, (X^n) converge in distribuzione alla legge di Benford e, per il Corollario 2.1, è anche Benford con probabilità uno.

□

Prodotti di variabili aleatorie indipendenti

Quando si analizza l'equazione (2.1) con fattori indipendenti $X_1, X_2, \dots$ è importante osservare che, nel contesto dell'indipendenza, la legge di Benford riveste un ruolo molto particolare per i prodotti finiti. Informalmente diremmo che "se uno dei fattori in un prodotto di fattori indipendenti segue la legge di Benford, allora l'intero prodotto seguirà esattamente tale legge, indipendentemente dagli altri fattori".

Teorema 2.8. *Siano X e Y variabili aleatorie indipendenti tali che $\mathbb{P}(XY = 0) = 0$. Allora:*

i. Se X è Benford lo è anche XY;

ii. Se $S(X)$ e $S(XY)$ hanno la stessa distribuzione allora o $\log S(Y)$ è razionale con probabilità uno oppure X è Benford.

Dimostrazione. (i) Segue direttamente dal Teorema 1.5 e dal Teorema 1.8.
(ii) Osserviamo innanzitutto che
$\log S(XY) = \langle \log S(X) + \log S(Y) \rangle$, infatti $\log S(XY) = \langle \log|XY| \rangle$, mentre

$$\langle \log S(X) + \log S(Y) \rangle = \langle \langle \log|X| \rangle + \langle \log|Y| \rangle \rangle = \langle \log|XY| \rangle.$$

Inoltre, le variabili aleatorie $X_0 = \log S(X)$ e $Y_0 = \log S(Y)$ sono indipendenti, questo poichè X_0 e Y_0 sono trasformazioni misurabili rispettivamente di X e Y e trasformazioni misurabili di variabili aleatorie indipendenti producono variabili aleatorie tra loro ancora indipendenti. Il Lemma 1.3 (iv) implica quindi che

$$\widehat{P_{\log S(XY)}} = \widehat{P_{\langle X_0+Y_0\rangle}} = \widehat{P_{X_0}} \cdot \widehat{P_{Y_0}}.$$

Se $S(X)$ e $S(XY)$ hanno la stessa distribuzione, per quanto appena visto ne seguirebbe che

$$\widehat{P_{X_0}}(k) \cdot \left(1 - \widehat{P_{Y_0}}(k)\right) = 0 \text{ per ogni } k \in \mathbb{Z}.$$

Abbiamo quindi due possibilità:

- $\widehat{P_{Y_0}}(k) \neq 1$ per ogni $k \neq 0$, allora per il Lemma 1.3 (vi) si avrà che $\widehat{P_{X_0}} = \widehat{\lambda_{0,1}}$, ovvero X è Benford;

- Altrimenti, $\widehat{P_{Y_0}}(k) = 1$ per un qualche $k_0 \neq 0$, ossia, per definizione, $\mathbb{E}[e^{2\pi i k_0 Y_0}] = 1$. Ciò comporta quindi che $e^{2\pi i k_0 Y_0} = 1$ quasi certamente e, affinché ciò avvenga, è necessario che $2\pi i k_0 Y_0$ sia un multiplo intero di $2\pi i$, ovvero $2\pi i k_0 Y_0 = 2\pi i n$ per un qualche $n \in \mathbb{Z}$.
  Ne segue quindi che $Y_0 = \frac{n}{k_0}$ per $n \in \mathbb{Z}$, ossia $Y_0 \in \frac{1}{k_0}\mathbb{Z}$ con probabilità uno. Pertanto $P_{Y_0}\left(\frac{1}{k_0}\mathbb{Z}\right) = 1$; dunque $k_0 Y_0 = k_0 \log S(Y)$ è un intero con probabilità 1, dunque $\log S(Y)$ è razionale con probabiltà 1.

$\square$

Ora che abbiamo chiarito il ruolo della legge di Benford nei prodotti finiti, possiamo passare ad esaminare le successioni di prodotti di variabili aleatorie indipendenti e identicamente distribuite, ossia considerare (2.1) con una sequenza i.i.d. (X_n). Come nel caso delle sequenze di potenze di una variabile aleatoria, utilizzando gli strumenti sviluppati nel Capitolo 1 tramite i coefficienti di Fourier, è semplice determinare se la successione di prodotti converge in distribuzione alla legge di Benford, o se segue la legge di Benford con probabilità uno.
Prima di entrare nel vivo, è utile però trattare due teoremi per le passeggiate aleatorie sulla circonferenza che ritorneranno nelle dimostrazioni del Teorema 2.11 e del Teorema 2.12. Il primo teorema stabilisce una condizione necessaria e sufficiente affinché la distribuzione dei cammini di una passeggiata aleatoria sul cerchio unitario converga verso una distribuzione uniforme,

mentre il secondo stabilisce una condizione necessaria e sufficiente sulla distribuzione sottostante agli incrementi affinché quasi tutti i cammini siano u.d. mod 1.

Teorema 2.9. *Per ogni* $n \in \mathbb{N}$, *sia* $S_n = X_1 + \cdots + X_n$, *dove* $X_1, X_2, \ldots$ *sono variabili aleatorie i.i.d. Le seguenti affermazioni sono equivalenti:*

i. $\lim_{n\to\infty} \mathbb{P}\left(\langle S_n \rangle \leq s\right) = s$ *per ogni* $0 \leq s < 1$*;*

ii. $\mathbb{P}\left(X_1 \in x + \frac{1}{m}\mathbb{Z}\right) < 1$ *per ogni* $x \in \mathbb{R}$ *e* $m \in \mathbb{N}$*.*

Dimostrazione. Supponiamo che valga (ii), allora per il Lemma 1.3 (ii), $|\widehat{P_{\langle X_1 \rangle}(k)}| < 1$ per ogni $k \in \mathbb{Z} \setminus \{0\}$. Pertanto, per il Lemma 1.3 (iv),

$$\begin{aligned} \widehat{P_{\langle S_n \rangle}}(k) &= \widehat{P_{\langle \sum_{j=1}^n X_j \rangle}}(k) \overset{X_1,\ldots,X_n \text{ i.i.d.}}{=} \prod_{j=1}^n \widehat{P_{\langle X_j \rangle}}(k) \\ &= \left(\widehat{P_{\langle X_1 \rangle}}(k)\right)^n \xrightarrow{n\to\infty} 0 \text{ per ogni } k \in \mathbb{Z} \setminus \{0\}, \end{aligned}$$

ovvero $\widehat{P_{\langle S_n \rangle}}(k) \xrightarrow{n\to\infty} \widehat{\lambda_{0,1}}(k)$ per ogni $k \in \mathbb{Z} \setminus \{0\}$, quindi per il Lemma 1.3 (v) si ha che vale (i). Supponiamo ora che non valga (ii), ovvero $\mathbb{P}\left(X_1 \in x + \frac{1}{m}\mathbb{Z}\right) = 1$ per un qualche $x \in \mathbb{R}$ e $m \in \mathbb{N}$. In questo caso allora,

$$\widehat{P_{\langle X_1 \rangle}}(k) = \mathbb{E}[e^{2\pi i m X_1}] = e^{2\pi i m x},$$

quindi, per ogni $n \in \mathbb{N}$,

$$|\widehat{P_{\langle S_n \rangle}}(m)| = |\widehat{P_{\langle \sum_{j=1}^n X_j \rangle}}(m)| = |\widehat{P_{\langle X_1 \rangle}}(m)^n| = |e^{2\pi i m n x}| = 1,$$

per cui chiaramente (i) non è soddisfatta.

$\square$

Teorema 2.10. *Per ogni* $n \in \mathbb{N}$, *sia* $S_n = X_1 + \cdots + X_n$, *dove* $X_1, X_2, \ldots$ *sono variabili aleatorie i.i.d. Le seguenti affermazioni sono equivalenti:*

i. La sequenza (S_n) *è u.d. mod 1 con probailità 1;*

ii. $\mathbb{P}\left(X_1 \in \frac{1}{m}\mathbb{Z}\right) < 1$ *per ogni* $m \in \mathbb{N}$*.*

Teorema 2.11. *Si considerino* $X_1, X_2, \ldots$ *variabili aleatorie i.i.d. con* $\mathbb{P}(X_1 = 0) = 0$*. Le seguenti affermazioni sono equivalenti:*

i. La successione $\left(\prod_{j=1}^n X_j\right)$ *converge in distribuzione alla legge di Benford;*

ii. $\mathbb{P}\left(\log|X_1| \in x + \frac{1}{m}\mathbb{Z}\right) < 1$ *per ogni* $x \in \mathbb{R}$ *e* $m \in \mathbb{N}$*;*

iii. $\mathbb{P}\left(S(X_1^m) = t\right) < 1$ *per ogni* $t \in [1, 10)$ *e* $m \in \mathbb{N}$*.*

Dimostrazione. Ponendo $Y_n = \log|X_n|$ per $n \in \mathbb{N}$, l'equivalenza di (i) e (ii) segue direttamente dal Teorema 1.5 e dal Teorema 2.9. L'equivalenza di (ii) e (iii) segue invece dalla formula (2.2).

□

Il prossimo risultato rappresenta una naturale generalizzazione, attraverso il caso speciale in cui $X_1 = \theta$ con probabilità uno, del Teorema di Weyl, il quale afferma che "$(n\theta)$ è u.d. mod 1 se e solo se θ è irrazionale".

Teorema 2.12. *Si considerino* $X_1, X_2, \ldots$ *variabili aleatorie i.i.d. con* $\mathbb{P}(X_1 = 0) = 0$*. Le seguenti affermazioni sono equivalenti:*

i. La sequenza $\left(\prod_{j=1}^{n} X_j\right)$ *è Benford con probabilità 1;*

ii. $\mathbb{P}\left(\log|X_1| \in \frac{1}{m}\mathbb{Z}\right) < 1$ *per ogni* $m \in \mathbb{N}$*;*

iii. $\mathbb{P}\left(S(X_1^m) = 1\right) < 1$ *per ogni* $m \in \mathbb{N}$*.*

Dimostrazione. Ponendo $Y_n = \log|X_n|$ per $n \in \mathbb{N}$, l'equivalenza di (i) e (ii) segue direttamente dal Teorema 1.5 e dal Teorema 2.10. L'equivalenza di (ii) e (iii) segue invece dalla formula (2.2).

□

Una sequenza $\left(\prod_{j=1}^{n} X_j\right)$ con fattori i.i.d. potrebbe essere Benford con probabilità uno ma non convergere in distribuzione alla legge di Benford. Vediamone un esempio.

Esempio 2.3

Poniamo $\mathbb{P}(X_n = 2) = 1$ per ogni $n \in \mathbb{N}$, allora $X_1, X_2 \ldots$ sono variabili aleatorie i.i.d., soddisfano (ii) del Teorema 2.12, pertanto $\left(\prod_{j=1}^{n} X_j\right)$ è Benford con probabilità 1.

Per quanto riguarda la condizione (ii) del Teorema 2.11, ovvero

$$\mathbb{P}\left(\log|X_1| \in x + \frac{1}{m}\mathbb{Z}\right) < 1 \text{ per ogni } x \in \mathbb{R} \text{ e } m \in \mathbb{N},$$

essa non è verificata, infatti, scegliendo $x = \log 2$, si ha che

$$\mathbb{P}\left(\log|X_1| \in \log 2 + \frac{1}{m}\mathbb{Z}\right) = 1 \text{ per ogni } m \in \mathbb{N}.$$

Di conseguenza, la successione $\left(\prod_{j=1}^{n} X_j\right) = (2^n)$ non converge in distribuzione alla legge di Benford.

Il seguente corollario garantisce però l'implicazione inversa.

Corollario 2.2. *Si considerino $X_1, X_2, \ldots$ variabili aleatorie i.i.d. e supponiamo che $\left(\prod_{j=1}^{n} X_j\right)$ converga in distribuzione alla legge di Benford. Allora è Benford con probabilità 1.*

Dimostrazione. E' sufficiente osservare che (iii) del Teorema 2.8 implica (iii) del Teorema 2.9.

□

Si osservi che il Teorema 2.11 (ii) e il Teorema 2.12 (ii) valgono a meno che la variabile aleatoria X_1 sia discreta, ovvero a meno che esista un insieme al più numerabile $C \subset \mathbb{R}$ tale che $\mathbb{P}(X_1 \in C) = 1$. Di conseguenza, i Teoremi 2.11 e 2.12 insieme implicano il seguente risultato utile.

Corollario 2.3. *Siano $X_1, X_2, \ldots$ variabili aleatorie i.i.d. Se X_1 non è discreta, allora $\left(\prod_{j=1}^{n} X_j\right)$ converge in distribuzione alla legge di Benford ed è Benford con probabilità 1.*

Si osservi però che, anche se X_1 è discreta, essa potrebbe comunque soddisfare la condizione (ii) del Teorema 2.11.

Esempio 2.4
Consideriamo una variabile aleatoria $X_1 > 0$ tale che $\mathbb{P}(X_1 = 2^j) = 2^{-j}$ per ogni $j \in \mathbb{N}$. Chiaramente, X_1 è discreta ($C = \{2^j : j \in \mathbb{N}\}$) e $\mathbb{P}\left(\log|X_1| = j \log 2\right) = 2^{-j}$. Tuttavia, $\log 2$ è irrazionale, di conseguenza, dati un qualsiasi $x \in \mathbb{R}$ e $m \in \mathbb{N}$, i due insiemi $\{j \log 2 : j \in \mathbb{N}\}$ e $x + \frac{1}{m}\mathbb{Z}$ hanno al più un elemento in comune, pertanto la probabilità che $\log|X_1|$ appartenga a $x + \frac{1}{m}\mathbb{Z}$ è limitata dalla probabilità di un singolo valore $j \log 2$ di verificarsi, la quale è massima (1/2) per $j = 1$ e decresce per $j > 1$. Si ha quindi che $\mathbb{P}\left(\log|X_1| \in x + \frac{1}{m}\mathbb{Z}\right) \leq 1/2$. Dunque è soddisfatta la condizione (ii) del Teorema 2.8, perciò $\left(\prod_{j=1}^{n} X_j\right)$ converge in distribuzione alla legge di Benford e, per il Corollario 2.3, è anche Benford con probabilità 1.

Un'altra conseguenza immediata dei Teoremi 2.6 e 2.12 è che le proprietà di essere Benford quasi certamente delle successioni (X_1^n) e $\left(\prod_{j=1}^{n} X_j\right)$ sono correlate.

Corollario 2.4. *Siano $X_1, X_2, \ldots$ variabili aleatorie i.i.d. Se (X_1^n) è Benford con probabilità uno allora lo è anche $\left(\prod_{j=1}^{n} X_j\right)$.*

Dimostrazione. Per il Teorema 2.6, (X_1^n) è Benford con probabilità uno se e solo se $\mathbb{P}(\log|X_1| \in \mathbb{Q}) = 0$, ovvero se e solo se $\log|X_1|$ assume quasi sicuramente valori irrazionali. Poiché $\mathbb{Q}$ è denso in $\mathbb{R}$, se $\log|X_1|$ non appartiene quasi certamente a $\mathbb{Q}$, allora la sua distribuzione non può concentrarsi su nessun insieme "reticolare" come $\frac{1}{m}\mathbb{Z}$, ovvero

$$\mathbb{P}\left(\log|X_1| \in x + \frac{1}{m}\mathbb{Z}\right) \neq 1 \text{ per ogni } m \in \mathbb{N}.$$

Ciò implica (ii) del Teorema 2.12, pertanto $\left(\prod_{j=1}^{n} X_j\right)$ è anch'essa Benford con probabilità 1.

□

In generale, non vale il viceversa del Corollario 2.4, come si osserva nel prossimo esempio.

Esempio 2.5
Sia $0 < p < 1$ e siano $X_1, X_2, \dots$ variabili aleatorie positive i.i.d. aventi stessa funzione di ripartizione

$$F_{X_1}(x) = \begin{cases} \frac{px}{\sqrt[3]{10}} \text{ se } 0 \leq x < \sqrt[3]{10}, \\ 1 \text{ se } x \geq \sqrt[3]{10}, \end{cases}$$

cioè X_1 è uniformemente distribuita su $(0, \sqrt[3]{10})$ con probabilità p e, altrimenti (ovvero con probabilità $1-p$) è uguale a $\sqrt[3]{10}$. Poiché

$$\begin{aligned}
\mathbb{E}[e^{6\pi in\log X_1}] &= \int_0^{\sqrt[3]{10}} \frac{p}{\sqrt[3]{10}} e^{6\pi in\log x}\, dx + (1-p)\cdot e^{6\pi in \log e \ln x} \\
&= \int_0^{\sqrt[3]{10}} \frac{p}{\sqrt[3]{10}} e^{6\pi in\log e \ln x}\, dx + (1-p)\cdot e^{2\pi in} \\
&= \int_0^{\sqrt[3]{10}} \frac{p}{\sqrt[3]{10}} x^{6\pi in\log e}\, dx + (1-p) \\
&= p\frac{(\sqrt[3]{10})^{6\pi in \log e}}{6\pi in\log e + 1} + (1-p) \\
&= p\frac{e^{2\pi in}}{6\pi in\log e + 1} + (1-p) \\
&= \frac{p}{6\pi in\log e + 1} + (1-p) \\
&= \frac{1 + i6\pi(1-p)n\log e}{1 + i6\pi n\log e} \\
&\xrightarrow{n\to\infty} 1 - p \neq 0,
\end{aligned}$$

per il Teorema 2.5 (X_1^n) non converge in distribuzione alla legge di Benford. Inoltre, in quanto $\mathbb{P}\left(\log X_1 = \frac{1}{3}\right) = \mathbb{P}\left(X_1 = \sqrt[3]{10}\right) = 1 - p > 0$, non è nemmeno Benford con probabilità 1.
D'altra parte, X_1 non è discreta, dunque per il Corollario 2.3, $\left(\prod_{j=1}^n X_j\right)$ converge in distribuzione alla legge di Benford ed è Benford con probabilità 1.

Infine, si noti che un caso speciale particolarmente importante del Corollario 2.3 si verifica quando X_1 ha una densità. In questo caso, $\mathbb{P}(X_1 \in C) = 0$ per ogni insieme numerabile $C \subset \mathbb{R}$, quindi X_1 non è chiaramente discreta.

Corollario 2.5. *Se le variabili aleatorie i.i.d $X_1, X_2, \ldots$ sono dotate di una densità, allora la sequenza $\left(\prod_{j=1}^n X_j\right)$ converge in distribuzione alla legge di Benford ed è Benford con probabilità 1.*

Richiamando il fatto (menzionato all'inizio di questo capitolo) che
$X_n \xrightarrow{\text{q.c.}} X$ implica $X_n \xrightarrow{D} X$, ma non viceversa, il lettore noterà che, per quanto riguarda le proprietà di Benford dell'equazione (2.1), questa implicazione sembra invertita, almeno nei due scenari studiati qui: per i Corollari 2.1 e 2.2, se la successione $\left(\prod_{j=1}^n X_j\right)$ converge in distribuzione alla legge di Benford, allora è Benford con probabilità uno, ma in generale non vale il viceversa.
Per mettere queste osservazioni in prospettiva, può essere utile notare che nel caso ancora più semplice di una successione i.i.d. $(X_1, X_2, \ldots)$ le due proprietà sono in realtà equivalenti.

Teorema 2.13. *Siano $X_1, X_2, \ldots$ variabili aleatorie i.i.d. Le seguenti affermazioni sono equivalenti*

i. La sequenza (X_n) è Benford con probabilità 1;

ii. La sequenza (X_n) converge in distribuzione alla legge di Benford;

iii. X_1 è Benford.

Dimostrazione. $(i) \Rightarrow (iii)$ Supponiamo che $(X_1, X_2, \ldots)$ sia Benford con probabilità 1. Questo vuol dire che per ogni funzione continua e limitata $f : [1, 10) \to \mathbb{R}$ si ha che

$$\frac{1}{N}\sum_{n=1}^{N} f\left(S(X_n)\right) \xrightarrow{q.c.} \int_1^{10} f(t)t^{-1}\log e\, dt \text{ per } N \to \infty.$$

Il Teorema della Convergenza Dominata implica che

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}\left[f\left(S(X_n)\right)\right] = \int_1^{10} f(t)t^{-1} \log e \, dt,$$

ma $\mathbb{E}\left[f\left(S(X_n)\right)\right] = \mathbb{E}\left[f\left(S(X_1)\right)\right]$ e dunque $\mathbb{E}\left[f\left(S(X_1)\right)\right] = \int_1^{10} f(t)t^{-1} \log e \, dt$. Poiché f era arbitraria, $\mathbb{P}(S(X_1) \leq t) = \log t \quad \forall t \in [1, 10)$, ovvero X_1 è Benford.
$(i) \Leftarrow (iii)$ Supponiamo ora che X_1 sia Benford e fissiamo un $t \in [1, 10)$. La Legge Forte dei Grandi Numeri applicata alla successione i.i.d. $\left(\mathbf{1}_{[1,t]}\left(S(X_n)\right)\right)$ implica che

$$\begin{aligned} \frac{\#\left\{1 \leq n \leq N : S(X_n) \leq t\right\}}{N} &= \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{[1,t]}\left(S(X_n)\right) \\ &\xrightarrow{q.c.} \mathbb{E}\left[\mathbf{1}_{[1,t]}\left(S(X_1)\right)\right] = \log t \qquad \text{per } N \to \infty. \end{aligned}$$

Ciò mostra che (X_n) è Benford con probabilità 1.
$(ii) \iff (iii)$ Segue direttamente dalla Definizione 2.1.

$\square$

Somme di variabili aleatorie

Ci concentreremo ora sullo studio della legge di Benford applicata alle somme di variabili aleatorie. Il comportamento statistico della funzione dei significandi per le somme è decisamente più complesso rispetto a quello che si osserva nei prodotti. La principale causa di questa complessità è che la funzione dei significandi della somma di due o più numeri dipende non solo dalla funzione dei significandi di ciascun numero (come accade nei prodotti), ma anche dai rispettivi esponenti. Ad esempio, si può osservare che

$$S(3 \cdot 10^3 + 2 \cdot 10^2) = 3.2, \quad S(3 \cdot 10^2 + 2 \cdot 10^2) = 5, \quad S(3 \cdot 10^2 + 2 \cdot 10^3) = 2.3$$

sono tutti diversi, mentre

$$S(3 \cdot 10^3 \times 2 \cdot 10^2) = S(3 \cdot 10^2 \times 2 \cdot 10^2) = S(3 \cdot 10^2 \times 2 \cdot 10^3) = 6.$$

Da un punto di vista pratico, questa difficoltà risiede nel fatto che, per qualsiasi numero reale positivo x e y, il valore di $\log(x + y)$ non può essere facilmente espresso in termini di $\log x$ e $\log y$, mentre $\log(xy) = \log x + \log y$, tuttavia tale fatto risulta rilevante per la conformità alla legge di Benford tramite il Teorema 1.5.

Ricordiamo dal Teorema 2.8 che la legge di Benford è estremamente attrattiva nel caso di moltiplicazione di variabili casuali indipendenti, poiché il prodotto di variabili positive segue la legge di Benford se almeno una delle singole variabili la segue. Al contrario, la somma di due variabili casuali indipendenti potrebbe non seguire la legge di Benford anche se entrambe le variabili la seguono.

Esempio 2.6
Siano X e Y variabili aleatorie i.i.d. Benford a valori in $[1, 10)$, ossia $\mathbb{P}(X \leq t) = \mathbb{P}(Y \leq t) = \log t$ per ogni $1 \leq t < 10$. Allora X+Y non è Benford. La densità di $S(X+Y)$ è data da

$$f_{S(X+Y)}(t) = \frac{2\log e}{t} \cdot \begin{cases} \log(10t-1) \text{ se } 1 \leq t < 1.1, \\ |\log(t-1)| \text{ se } 1.1 \leq t < 10, \end{cases}$$

e quindi

$$\begin{aligned} \max_{1\leq d\leq 9} |\mathbb{P}(D_1(X+Y)=d) - \log(1+d^{-1})| &\geq \log e \cdot \max_{1\leq d\leq 9} \left| \int_{d-1}^{d} \frac{2\log t - 1}{t+1}\, dt \right| \\ &= \log e \cdot \int_1^2 \frac{1-2\log t}{t+1}\, dt \\ &= 0.125. \end{aligned}$$

In analogia con (2.1) si consideri ora

$$\left(\sum_{j=1}^{n} X_j\right) = (X_1, X_1+X_2, \dots), \text{ con } X_1, X_2, \dots \text{identicamente distribuite} \tag{2.3}$$

dove, come per i prodotti, verranno discussi in dettaglio solo gli scenari di termini X_j identici e indipendenti, rispettivamente. Si assuma innanzitutto che tutti i termini $X_1, X_2, \dots$ in (2.3) siano uguali, allora la successione in (2.3) assume semplicemente la forma di (nX), dove X è una qualsiasi variabile aleatoria a valori reali. Introduciamo il seguente teorema che ricorrerà nella dimostrazione del Teorema 2.15.

Teorema 2.14. *Sia X una variabile aleatoria Benford. Allora per ogni $a \in \mathbb{R}$ e $k \in \mathbb{Z}$ tali che $ak \neq 0$, la variabile aleatoria aX^k è anch'essa Benford.*

Il prossimo teorema è l'analogo per le somme dei Teoremi 2.5 e 2.6.

Teorema 2.15. *Sia X una variabile aleatoria. Allora:*

i. La successione (nX) converge in distribuzione alla legge di Benford se e solo se X è Benford;

ii. *Se X non è Benford non vi è alcuna sttosuccessione di (nX) che converga in distribuzione alla legge di Benford;*

iii. *Con probabilità 1, (nX) non è una successione Benford .*

Dimostrazione. Si assuma che $\mathbb{P}(X = 0) = 0$. Per dimostrare (i), poniamo $Y_n = \langle \log|nX| \rangle$ e si osservi che per ogni $n \in \mathbb{N}$ e $k \in \mathbb{Z}$,

$$\begin{aligned}|\widehat{P_{Y_n}}(k)| &= \left|\mathbb{E}[e^{2\pi ik\log|nX|}]\right| \\ &= \left|e^{2\pi ik\log n}\mathbb{E}[e^{2\pi ik\log|X|}]\right| \\ &\overset{|e^{2\pi ik\log n}|=1}{=} \left|\widehat{P_{\langle\log|X|\rangle}}(k)\right|.\end{aligned}$$

Per il Teorema 2.14, se X è Benford allora lo è anche nX per ogni n.
Per vedere (ii), si osservi che se X non è Benford allora $\widehat{P_{\langle\log|X|\rangle}}(k) \neq 0$ per qualche $k \neq 0$, questo poiché X è Benford se e solo se $\log|X|$ è u.d. mod 1 (Teorema 1.5), che equivale a richiedere $\langle \log|X| \rangle$ sia uniforme su (0,1). Dunque, per il Lemma 1.3(vi), X è Benford se e solo se

$$\widehat{P_{\langle\log|X|\rangle}}(k) = \begin{cases} 1 & \text{se } k = 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Di conseguenza, $|\widehat{P_{Y_n}}(k)| = \left|\widehat{P_{\langle\log|X|\rangle}}(k)\right| > 0$ per ogni n, ne segue che nessuna sottosuccessione di (Y_n) converge in distribuzione a $U(0,1)$, ovvero nessuna sottosuccessione di (nX) converge in distribuzione alla legge di Benford.
Se $P(X = 0) = \delta > 0$, allora $P(nX = 0) = \delta$ per ogni $n \in \mathbb{N}$, da cui risulta chiaro che (i) e (ii) valgono anche in questo caso.
Resta da dimostrare (iii). Si ha che la successione $(\log n)$ non è u.d. mod 1, questo poiché la succesione $\left(\frac{1}{N}\#\left\{1 \leq n \leq N : \langle \log n \rangle \leq s\right\}\right)_{N\in\mathbb{N}}$ ha rispettivamente lim inf e lim sup pari a $\frac{1}{9}(10^s - 1)$ e $\frac{10}{9}(1 - 10^{-s})$. Per il Teorema 1.5 pertanto (n) non è Benford e, per il Teorema 2.11, nemmeno (nx) è Benford per ogni $x \in \mathbb{R}$.
Da ciò segue immediatamente che

$$P\left(\{\omega : nX(\omega) \text{ è una sequenza di Benford}\}\right) = P(\emptyset) = 0.$$

□

Consideriamo ora in (2.3) $X_1, X_2, \ldots$ variabili aleatorie i.i.d. aventi tutte distribuzioni diverse a meno che X_1 sia identicamente 0 con probabilità 1. Ciò che afferma il seguente teorema, che è l'analogo per le somme dei Teoremi 2.11 e 2.12, potrebbe non sorprendere per quanto visto fino ad ora.

Teorema 2.16. *Siano $X_1, X_2, \dots$ variabili aleatorie i.i.d. aventi momenti secondi finiti, ossia $\mathbb{E}[X_1^2] < \infty$. Allora*

i. Nessuna sottosuccessione di $\left(\sum_{j=1}^n X_j\right)$ converge in distribuzione alla legge di Benford;

ii. La successione $\left(\sum_{j=1}^n X_j\right)$ non è Benford con probabilità 1.

Dimostrazione. Supponiamo che $\mathbb{E}[X_1] \neq 0$ e dimostriamo che (i) e (ii) sono verificate.
Innanzitutto vogliamo dimostrare che nessuna sottosuccessione di $\left(\sum_{j=1}^n X_j\right)$ converge in distribuzione alla legge di Benford, ovvero che nessuna sottosuccessione di $\left(S\left(\sum_{j=1}^n X_j\right)\right)$ converge a $S(X)$ dove X è una certa variabile aleatoria Benford.
Poiché per ipotesi $X_1, X_2, \dots$ sono i.i.d. e $\mathbb{E}[X_1^2] < \infty$, per la Legge Forte dei Grandi Numeri si ha che $\left(\frac{1}{n}\left|\sum_{j=1}^n X_j\right|\right)$ converge quasi certamente, e di conseguenza in distribuzione, a $|\mathbb{E}[X_1]|$. Di conseguenza si ha che $\mathbb{P}\left(\sum_{j=1}^n X_j = 0\right) \to 0$ per $n \to \infty$, infatti

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left(\sum_{j=1}^n X_j = 0\right) &= \lim_{n\to\infty} \mathbb{P}\left(\sum_{j=1}^n X_j \leq 0, \sum_{j=1}^n X_j \geq 0\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\left|\sum_{j=1}^n X_j\right| \leq 0\right) \\
&= \mathbb{P}\left(|\mathbb{E}[X_1]| \leq 0\right) \\
&\overset{\mathbb{E}[X_1]\neq 0}{=} 0.
\end{aligned}$$

Osserviamo inoltre che

$$\begin{aligned}
\log S\left(\sum_{j=1}^n X_j\right) &= \log\left(10^{\log\left|\sum_{j=1}^n X_j\right| - \left\lfloor \log\left|\sum_{j=1}^n X_j\right|\right\rfloor}\right) \\
&= \log\left|\sum_{j=1}^n X_j\right| - \left\lfloor \log\left|\sum_{j=1}^n X_j\right|\right\rfloor \\
&= \left\langle \log\left|\sum_{j=1}^n X_j\right| \right\rangle \\
&= \left\langle \log\left(\frac{1}{n}\left|\sum_{j=1}^n X_j\right|\right) + \log n \right\rangle.
\end{aligned}$$

Da queste due osservazioni concludiamo che ogni qualvolta $\sum_{j=1}^n X_j \neq 0$, ogni sottosuccessione di $\left(S\left(\frac{1}{n}\sum_{j=1}^n X_j\right)\right)$ o non converge in distribuzione oppure converge ad una costante: in nessun caso, pertanto, il limite è una variabile aleatoria Benford, perciò è stato dimostrato quanto voluto.
Vogliamo ora provare che la successione $\left(\sum_{j=1}^n X_j\right)$ non è Benford con probabilità 1.
Poiché $\left|\sum_{j=1}^n X_j\right| \to \infty$ con probabilità 1, segue da

$$\log\left|\sum_{j=1}^n X_j\right| - \log n = \log\frac{1}{n}\left|\sum_{j=1}^n X_j\right| \xrightarrow{q.c.} |\mathbb{E}[X_1]|,$$

e dalle Proposizioni 1.2 (i) e 1.3 che la successione $\left(\sum_{j=1}^n X_j\right)$ non è Benford con probabilità 1.
Rimane ora da considerare il caso $\mathbb{E}[X_1] = 0$, senza perdere di generalità, possiamo assumere che $\mathbb{E}[X_1^2] = 1$, pertanto $Var[X_1] = 1$.
Per il Teorema del Limite Centrale si ha che $\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n X_j\right)$ converge in distribuzione ad una variabile aleatoria con legge gaussiana con media 0 e varianza 1, ovvero converge in distribuzione alla distribuzione normale standard. Dunque, per ogni n sufficientemente grande e a meno di una rotazione (cioè, un'aggiunta modulo 1) di $[0,1)$, la distribuzione di $\left\langle\log\left|\sum_{j=1}^n X_j\right|\right\rangle$ è arbitrariamente vicina alla distribuzione della variabile casuale $Y := \langle\log|Z|\rangle$, dove Z è una normale standard. Mostriamo ora che $P_y \neq \lambda_{0,1}$, ovvero Y non è uniforme in $[0,1)$. In primo luogo si noti che

$$F_Y(s) = 2\sum_{k\in Z}\left(\Phi(10^{s+k}) - \Phi(10^k)\right), \quad 0 \leq s < 1, \tag{2.4}$$

dove con $\Phi = (F_Z)$ denotiamo la funzione di distribuzione di una normale standard, ovvero

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

Questo poiché

$$\begin{aligned} F_Y(s) &= \mathbb{P}(Y \leq s) \\ &= \mathbb{P}(\langle\log|Z|\rangle \leq s) \\ &\overset{\text{dimostrazione Teo. 1.5}}{=} \mathbb{P}(S(Z) \leq 10^s) \\ &\overset{Z\sim\mathcal{N}(0,1)}{=} 2\sum_{k\in Z}\left(\Phi(10^k\cdot 10^s) - \Phi(10^k)\right), \quad 0 \leq s < 1. \end{aligned}$$

Si osservi inoltre che

$$\begin{aligned} |F_Y(s) - s| &\geq F_Y(s) - s \\ &> 2\left(\Phi(10^{s+0}) - \Phi(10^0)\right) - s \\ &= 2\left(\Phi(10^s) - \Phi(1)\right) - s \\ &=: R(s), \quad 0 \leq s < 1 \end{aligned}$$

e studiando la funzione R notiamo che

- R è concava in $[0, 1)$, infatti
$$R'(s) = 2\varphi(10^s)10^s \ln(10) - 1,$$
dove $\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, e
$$R''(s) = 2 \cdot 10^s \ln 10 \cdot \varphi(10^s)(1 - 10^s) < 0 \qquad \text{per } 0 \leq s < 1;$$
- $R(0) = 0$;
- $R'(0) = \frac{2}{\sqrt{2\pi e}} \ln(10) - 1 > 0$,

perciò

$$\max_{0 \leq s 1} |F_Y(s) - s| > \sup_{0 \leq s < 1} R(s) > 0.$$

Ciò mi dice pertanto che $P_y \neq \lambda_{0,1}$; di conseguenza, la successione $\left(\sum_{j=1}^n X_j\right)$, come ogni sua sottosuccessione, non converge in distribuzione alla legge di Benford.

Rimane ora da verificare (ii) nel caso $\mathbb{E}[X_1] = 0$ e per farlo utilizziamo una versione quasi certa del Teorema del Limite Centrale[8].

Poniamo

$$A = \left\{ w \in \Omega : \left(\sum_{j=1}^{n} X_j(w) \right) \text{ è una successione Benford} \right\},$$

avendo che le variabili aleatorie $X_1, X_2, \ldots$ sono definite sullo spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$.

Definiamo la successione $(x_n(w))$ con

$$x_n(w) = \log \frac{1}{\sqrt{n}} \left| \sum_{j=1}^{n} X_j(w) \right|, \quad n \in \mathbb{N}.$$

Per il Teorema 1.5 sappiamo che essa è u.d mod 1 se e solo se $\frac{1}{\sqrt{n}} \sum_{j=1}^n X_j(w)$ è Benford; pertanto $(x_n(w))$ è u.d. mod 1 per ogni $w \in A$ (per definizione di A). Per ogni intervallo $[a, b) \subset [0, 1)$ quindi (si veda l'Osservazione 1.3)

$$\frac{1}{\ln N} \sum_{n=1}^{N} \frac{\mathbf{1}_{[a,b)}\left(\langle x_n(w) \rangle\right)}{n} \to b - a \quad \text{per } N \to \infty.$$

Come conseguenza di [6, Teorema 2], per ogni $[a, b) \subset [0, 1)$ si ha che

$$\frac{1}{\ln N} \sum_{n=1}^{N} \frac{\mathbf{1}_{[a,b)}\left(\langle x_n \rangle\right)}{n} \xrightarrow{q.c.} F_Y(b) - F_Y(a),$$

con F_Y dato da (2.4). Precedentemente avevamo mostrato che $F_Y(s) \not\equiv s$ e dunque $\mathbb{P}(A) = 0$. In altre parole,
$\mathbb{P}\left(\left(\sum_{j=1}^{n} X_j(w)\right) \text{ è una successione Benford}\right) = 0.$

$\square$

2.3 Misture di distribuzioni

Come è già stato detto in precedenza, la legge di Benford appare in molti insiemi di dati reali, a questo punto però una domanda sorge spontanea: cosa hanno in comune i dati della popolazione di tremila contee statunitensi secondo il censimento del 2010 con l'uso delle tavole logaritmiche nel 1880, i dati numerici tratti dagli articoli di giornale degli anni '30 raccolti da Benford o le costanti fisiche universali esaminate da Knuth negli anni '60? Perché questi dati dovrebbero mostrare una distribuzione logaritmica delle cifre significative?
In realtà, la maggior parte dei dataset non segue rigorosamente la legge di Benford. Benford aveva già osservato che, mentre alcune delle sue tabelle si conformavano alla legge logaritmica in modo ragionevolmente accurato, molte altre no. Ma, come sottolinea Raimi, "ciò che si avvicinava di più, tuttavia, era l'unione di tutte le sue tabelle" ([12, p. 522]): combinando le tabelle dei pesi molecolari con le statistiche del baseball e le aree di drenaggio dei fiumi, si otteneva un risultato molto convincente.
Le prime spiegazioni della legge di Benford si basavano sull'ipotesi dell'esistenza di una sorta di tabella universale di costanti: Raimi[13] la visualizzava come un "deposito di dati tabulari nelle biblioteche del mondo", Knuth [7] come il "set immaginato di numeri". Gli studiosi che seguivano tale filone tentarono poi di dimostrare perché certi specifici insiemi di osservazioni reali fossero rappresentativi o di questa misteriosa tabella universale o dell'insieme di tutti i numeri reali.
L'approccio che lo stesso Benford seguì nel suo studio, e che appare più naturale, è tuttavia quello di pensare ai dati come provenienti da molte distribuzioni diverse. Dopotutto, egli aveva fatto uno sforzo "per raccogliere dati da quanti più campi possibile e includere una grande varietà di tipologie", notando che "la varietà di argomenti studiati e tabulati era la più ampia che il tempo e le energie permettessero."

L'obiettivo principale di questa sezione è fornire una derivazione statistica della legge di Benford sotto forma di un teorema simile al Teorema del Limite Centrale, che afferma che, se si estraggono campioni casuali da distribuzioni diverse e si combinano i risultati, allora — a condizione che il campionamento sia indipendente dalla scala o dalla base — i campioni combinati risultanti convergeranno alla distribuzione di Benford.
Si denoti con $\mathcal{M}$ l'insieme di tutte le misure di probabilità su $(\mathbb{R}^+, \mathcal{B}^+)$. Una misura di probabilità random (positiva di Borel) è una funzione $P : \Omega \to \mathcal{M}$ definita su un qualche spazio di probabilità sottostante $(\Omega, \mathcal{A}, \mathbb{P})$ tale che per ogni $B \in \mathcal{B}^+$ la funzione $w \mapsto P(w)(B)$ è una variabile aleatoria e, per ogni $w \in \Omega$, $P(w)$ è una misura di probabilità su $(\mathbb{R}^+, \mathcal{B}^+)$, dunque per ogni $a, b \in \mathbb{R}$ e per ogni boreliano $B \subset \mathbb{R}^+$,

$$\{w : a \leq P(w)(B) \leq b\} \in \mathcal{A}.$$

In una forma più astratta e concettuale, possiamo interpretare una misura di probabilità random (r.p.m) attraverso la seguente costruzione: si doti l'insieme $\mathcal{M}$ dellla topologia della convergenza in distribuzione, ovvero della più debole topologia che rende continue tutte le applicazioni del tipo

$$\mu \mapsto \int_{\mathcal{C}} f \, d\mu$$

per ogni funzione $f : \mathcal{C} \to \mathbb{R}$ continua e limitata. Con questa topologia $\mathcal{M}$ è uno spazio metrizzabile completo e separabile, possiamo pertanto definire la σ-algebra dei Boreliani di $\mathcal{M}$ $(\mathcal{B}_{\mathcal{M}})$, ovvero la più piccola σ-algebra contenente tutti i sottoinsiemi aperti di $\mathcal{M}$. Allora $\mathbb{P} \circ P^{-1}$ è una misura di probabilità su $(\mathcal{M}, \mathcal{B}_{\mathcal{M}})$. Vediamo due esempi concreti di r.p.m.

Esempio 2.7
Sia P una r.p.m. che ha distribuzione uniforme su (0,1) $(U(0,1))$ con probabilità $\frac{1}{2}$, altrimenti è esponenziale con media 1 $(Exp(1))$.
Per realizzare nella pratica P basta pensare al lancio di una moneta equa: se esce testa, P è una distribuzione uniforme su $(0, 1)$, mentre se esce croce è esponenziale con media 1.

Esempio 2.8
In questo esempio si vedrà una variante di una costruzione classica di L. Dubins e D. Freedman che, come vedremo successivamente, è una r.p.m. che porta alla legge di Benford.
Sia P una r.p.m. a supporto in $[1, 10)$, ovvero $P([1, 10)) = 1$ con probabilità 1, definita dalla sua funzione di ripartizione (random)

$$F_P(t) = F_{P(w)}(t) = P(w)([1, t]), \quad 1 \leq t < 10$$

come segue:

- si ponga $F_P(1) = 0$ e $F_P(10) = 1$;
- si fissi $F_P(10^{1/2})$ secondo la distribuzione uniforme su $[0, 1)$;
- si scelgano in modo indipendente $F_P(10^{1/4})$ e $F_P(10^{3/4})$ rispettivamente in modo uniforme in $\big[0, F_P(10^{1/2})\big)$ e $\big[F_P(10^{1/2}), 1\big)$;
- si proceda in questo modo.

Questa costruzione è nota per generare una r.p.m. quasi certamente [6, Lemma 9.28] e può essere facilmente verificato che è densa nell'insieme di tutte le misure di probabilità su $([1, 10), \mathcal{B}[1, 10))$, cioè genera misure di probabilità che sono arbitrariamente vicine ad una qualsiasi misura di probabilità di Borel su [1,10).

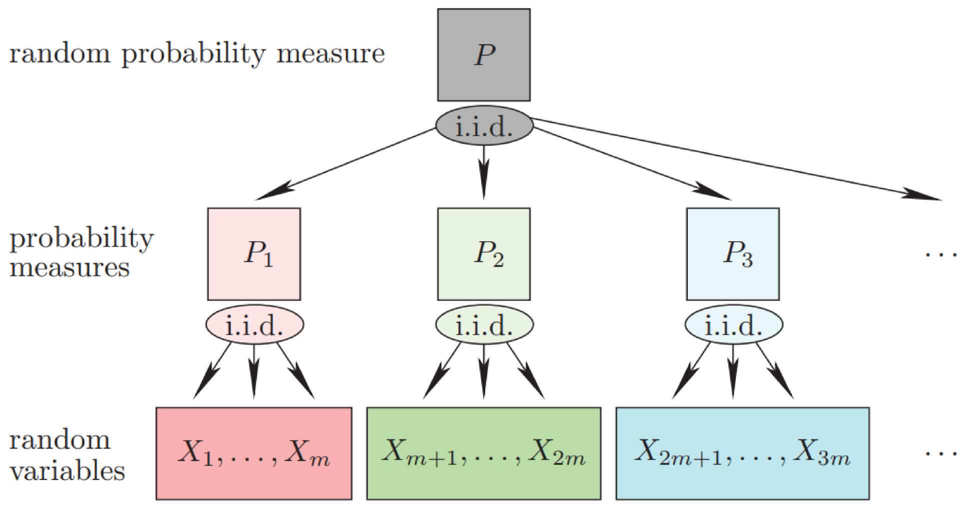
La prossima definizione formalizza il concetto di combinare dati provenienti da diverse distribuzioni. Essenzialmente, imita ciò che Benford fece combinando le statistiche del baseball con tabelle delle radici quadrate e dati presi dai giornali. L'idea è quella di usare una misura di probabilità casuale per generare una successione casuale di distribuzioni di probabilità e poi selezionare da ciascuna di queste le distribuzioni dei campioni casuali.

Definizione 2.2. *Sia m un intero positivo e P una r.p.m. Una successione di m-campioni P-casuali è una successione $(X_1, X_2, \dots)$ di variabili aleatorie su $(\Omega, \mathcal{A}, \mathbb{P})$ tali che $\forall j \in \mathbb{N}$ e per qualche successione di r.p.m. $(P_1, P_2, \dots)$ con $P_1 = P$ valgono le seguenti proprietà:*

i. Data $P_j = Q$, le variabili aleatorie $X_{(j-1)m+1}, X_{(j-1)m+2}, \dots, X_{jm}$ sono i.i.d. con distribuzione Q;

ii. Le variabili aleatorie $X_{(j-1)m+1}, X_{(j-1)m+2}, \dots, X_{jm}$ sono indipendenti da $P_i, X_{(i-1)m+1}, X_{(i-1)m+2}, \dots, X_{im}$ per ogni $i \neq j$.

In altre parole, data una qualsiasi sequenza $(X_1, X_2, \dots)$ di m-campioni P-casuali, per ogni $\omega \in \Omega$, le prime m variabili casuali $X_1, X_2, \dots, X_m$ sono un campione casuale (cioè i.i.d.) generato dalla distribuzione di probabilità $P_1(\omega)$, una distribuzione di probabilità casuale scelta secondo la r.p.m. P; la secondo m-tupla di variabili casuali $X_{m+1}, \dots, X_{2m}$ è un campione casuale generato da $P_2(\omega)$; e così via. Notiamo quindi che vi sono due livelli di casualità: prima una probabilità viene selezionata casualmente, poi un campione casuale viene estratto da questa distribuzione.
Se $(X_1, X_2, \dots)$ è una sequenza di m-campioni P-casuali per qualche m e qualche r.p.m. P, allora gli X_n sono, con probabilità 1, distribuiti in modo identico con una distribuzione che è la distribuzione media (attesa) di P (vedi la Proposizione 2.1 qui sotto), ma in generale non sono indipendenti. D'altra parte, dati $(P_1, P_2, \dots)$, gli X_n sono, con probabilità 1, indipendenti,



Genesi di una sequenza $(X_1, X_2, \dots)$ di m-campioni P-casuali

ma chiaramente non sono in generale distribuiti in modo identico.
Sebbene le sequenze di m-campioni P-casuali abbiano una struttura abbastanza semplice, non rientrano in nessuna delle categorie familiari di successioni di variabili aleatorie. Ad esempio, in generale non sono indipendenti, scambiabili, di Markov, martingala o successioni stazionarie.

Esempio 2.9
Supponiamo che la misura di probabilità random P sia tale che
$\mathbb{P}(P = \delta_1) = \mathbb{P}(P = \frac{1}{2}(\delta_1 + \delta_2)) = \frac{1}{2}$, dove δ_i con $i = 1, 2$ è la misura di Dirac concentrata rispettivamente in 1 e in 2. Osserviamo che $P = \frac{1}{2}(\delta_1 + \delta_2)$ mi dice che P è una distribuzione uniforme discreta su 1 e 2.
Si consideri una successione di 3-campioni P-casuali $(X_1, X_2, \dots)$. Allora le variabili aleatorie $X_1, X_2, \dots$

- non sono indipendenti: $\mathbb{P}(X_2 = 2) \neq \mathbb{P}(X_2 = 2|X_1 = 2)$, infatti

$$\begin{aligned}\mathbb{P}(X_2 = 2) &= \mathbb{P}(X_2 = 2|P = \delta_1)\mathbb{P}(P = \delta_1) \\ &+ \mathbb{P}(X_2 = 2|P = \frac{1}{2}(\delta_1 + \delta_2))\mathbb{P}(P = \frac{1}{2}(\delta_1 + \delta_2)) \\ &= 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{4}\end{aligned}$$

  e $\mathbb{P}(X_2 = 2|X_1 = 2) = \frac{1}{2}$;

- non sono scambiabili:
  $\mathbb{P}\left((X_1, X_2, X_3, X_4) = (1,1,1,2)\right) = \frac{9}{64} \neq \frac{3}{64} = \mathbb{P}\left((X_1, X_2, X_3, X_4) = (2,1,1,1)\right)$;
- non sono Markov: $\mathbb{P}(X_3 = 1 | X_1 = X_2 = 1) \neq \mathbb{P}(X_3 = 1 | X_2 = 1)$, infatti

$$\begin{aligned}
\mathbb{P}(X_3 = 1 | X_1 = X_2 = 1) &= \mathbb{P}(X_3 = 1 | X_2 = X_1 = 1, P = \delta_1)\mathbb{P}(P = \delta_1 | X_2 = X_1 = 1) \\
&\quad + \mathbb{P}\left(X_3 = 1 | X_2 = X_1 = 1, P = \frac{1}{2}(\delta_1 + \delta_2)\right) \\
&\quad \cdot \mathbb{P}\left(P = \frac{1}{2}(\delta_1 + \delta_2) | X_2 = X_1 = 1\right) \\
&= \frac{1}{2} \cdot \mathbb{P}(P = \delta_1 + | X_2 = X_1 = 1) + \frac{1}{2} \\
&= \frac{\mathbb{P}(X_2 = X_1 = 1 | P = \delta_1)\mathbb{P}(P = \delta_1)}{2 \cdot \mathbb{P}(X_2 = X_1 = 1)} + \frac{1}{2} \\
&= \frac{1 \cdot \frac{1}{2}}{2 \cdot \frac{5}{8}} + \frac{1}{2} \\
&= \frac{4}{10} + \frac{1}{2} \\
&= \frac{9}{10}
\end{aligned}$$

  e

$$\begin{aligned}
\mathbb{P}(X_3 = 1 | X_2 = 1) &= \mathbb{P}(X_3 = 1 | X_2 = 1, P = \delta_1)\mathbb{P}(P = \delta_1 | X_2 = 1) \\
&\quad + \mathbb{P}\left(X_3 = 1 | X_2 = 1, P = \frac{1}{2}(\delta_1 + \delta_2)\right)\mathbb{P}\left(P = \frac{1}{2}(\delta_1 + \delta_2) | X_2 = 1\right) \\
&= \mathbb{P}(P = \delta_1 | X_2 = 1) + \frac{1}{2} \cdot \mathbb{P}\left(P = \frac{1}{2}(\delta_1 + \delta_2) | X_2 = 1\right) \\
&= \mathbb{P}(P = \delta_1 | X_2 = 1) + \frac{1}{2} \cdot (1 - \mathbb{P}(P = \delta_1 | X_2 = 1)) \\
&= \frac{1}{2} \cdot \mathbb{P}(P = \delta_1 | X_2 = 1) + \frac{1}{2} \\
&= \frac{\mathbb{P}(X_2 = 1 | P = \delta_1)\mathbb{P}(P = \delta_1)}{2 \cdot \mathbb{P}(X_2 = 1)} + \frac{1}{2} \\
&= \frac{1}{3} + \frac{1}{2} \\
&= \frac{5}{6};
\end{aligned}$$

- non sono stazionarie: $\mathbb{P}\left((X_1, X_2, X_3) = (1,1,1)\right) \neq \mathbb{P}\left((X_2, X_3, X_4) = (1,1,1)\right)$, questo perché

$$\begin{aligned}
\mathbb{P}\left((X_1,X_2,X_3)=(1,1,1)\right) &= \mathbb{P}\left((X_1,X_2,X_3)=(1,1,1)|P=\delta_1\right)\cdot\frac{1}{2}\\
&+\mathbb{P}\left((X_1,X_2,X_3)=(1,1,1)|P=\frac{1}{2}\left(\delta_1+\delta_2\right)\right)\cdot\frac{1}{2}\\
&=1\cdot\frac{1}{2}+\frac{1}{8}\cdot\frac{1}{2}\\
&=\frac{9}{16}
\end{aligned}$$

e

$$\mathbb{P}\left((X_2,X_3,X_4)=(1,1,1)\right)=\frac{15}{32};$$

- non sono martingala: $\mathbb{E}[X_2|X_1=2]\neq\mathbb{E}[X_2]$ poiché

$$\begin{aligned}
\mathbb{E}[X_2|X_1=2] &= 1\cdot\mathbb{P}[X_2=1|X_1=2]+2\cdot\mathbb{P}[X_2=2|X_1=2]\\
&=\frac{1}{2}+2\cdot\frac{1}{2}\\
&=\frac{3}{2}
\end{aligned}$$

mentre

$$\begin{aligned}
\mathbb{E}[X_2] &= 1\cdot\mathbb{P}[X_2=1]+2\cdot\mathbb{P}[X_2]\\
&=\mathbb{P}(X_2=1|P=\delta_1)\mathbb{P}(P=\delta_1)\\
&+\mathbb{P}\left(X_2=1|P=\frac{1}{2}(\delta_1+\delta_2)\right)\mathbb{P}\left(P=\frac{1}{2}(\delta_1+\delta_2)\right)+2\cdot\frac{1}{4}\\
&=\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}+\frac{2}{4}\\
&=\frac{5}{4}.
\end{aligned}$$

Si ricordi che, data una r.p.m. P e un boreliano B, $P(B)$ è una variabile aleatoria con valori compresi tra 0 e 1. La seguente proprietà del valore atteso di $P(B)$, come funzione di B, è facile da verificare.

Proposizione 2.1. *Sia P una r.p.m. Allora $\mathbb{E}P$, definita come*

$$(\mathbb{E}P)(B)=\mathbb{E}[P(B)]=\int_\Omega P(\omega)(B)\,d\mathbb{P}(\omega)\quad\forall B\in\mathcal{B}^+,$$

è una misura di probabilità su $(\mathbb{R}^+,\mathcal{B}^+)$.

Passare da una r.p.m. P al suo valore atteso $\mathbb{E}P$ è uno step di regolarizzazione e poiché

$$\begin{aligned}|F_{\mathbb{E}P\circ S^{-1}}(t) - \log t| &= \left|\int_{\Omega} \left(F_{P(w)\circ S^{-1}}(t) - \log t\right)\, d\mathbb{P}(\omega)\right| \\ &\leq \sup_{\omega\in\Omega} \left|F_{P(w)\circ S^{-1}}(t) - \log t\right|, \quad 1 \leq t < 10,\end{aligned}$$

è chiaro che $\mathbb{E}P$ è vicino ad essere Benford almeno tanto quanto qualche $P(\omega)$. In realtà, come mostra il prossimo esempio, $\mathbb{E}P$ potrebbe addirittura conformarsi meglio alla legge di Benford di quanto non lo faccia alcun $P(w)$.

Esempio 2.10
(i) Si consideri una variabile aleatoria $X_T = U(0,T)$, dove $T \in [1,10)$ è a sua volta una variabile aleatoria avente densità

$$f_T(t) = \frac{2}{81}(10-t), \quad 1 \leq t < 10.$$

Per il Teorema 1.4 si ha che $\max_{t\in[1,10)} \left|F_{S(X_T)}(t) - \log t\right| \geq 0.1344$ per ogni valore di T. D'altra parte, per ogni $x > 0$

$$\begin{aligned}F_{\mathbb{E}X_T}(x) &= \mathbb{P}(\mathbb{E}X_T \leq x) \\ &= \int_1^{10} \mathbb{P}(X_t \leq x) f_T(t)\, dt\end{aligned}$$

e, poiché $X_t \sim U(0,t)$,

$$\mathbb{P}(X_t \leq x) = \begin{cases} 0 & \text{se } x \leq 0, \\ \frac{x}{t} & \text{se } 0 < x < t, \\ 1 & \text{se } x \geq t, \end{cases}$$

si ha che

$$\begin{aligned}F_{\mathbb{E}X_T}(x) &= \int_1^{10} \min\left\{\frac{x}{t}, 1\right\} \cdot \frac{2}{81}\,(10-t)\; dt \\ &= \begin{cases} \left(\frac{20}{81}\,(\log e)^{-1} - \frac{2}{9}\right) x & \text{se } 0 < x < 1, \\ \frac{1}{81}x^2 - \frac{20}{81}x\ln x + \frac{20}{81}\,(\log e)^{-1}\, x - \frac{19}{81} & \text{se } 1 \leq x < 10, \\ 1 & \text{se } x \geq 10. \end{cases}\end{aligned}$$

Di conseguenza, per $1 \leq t < 10$,

$$F_{S(\mathbb{E}X_T)}(t) = \sum_{k=0}^{\infty} \left(F_{\mathbb{E}X_T}\left(t10^{-k}\right) - F_{\mathbb{E}X_T}\left(10^{-k}\right)\right)$$
$$= \frac{1}{81}\left(t^2 - 20t\ln t + \left(\frac{200}{9}(\log e)^{-1} - 2\right)t + 1 - \frac{200}{9}(\log e)^{-1}\right).$$

Si osserva che $\max_{t\in[1,10)} \left|F_{S(\mathbb{E}X_T)}(t) - \log t\right| = 0.03847$, che è decisamente minore rispetto al corrispondente valore per ogni X_T.

(ii) In questo esempio vedremo che è possibile che $\mathbb{E}P$ sia Benford anche se ogni singola misura di probabilità $P(w)$ non lo è.

Consideriamo la mappa $g : [1, 10] \to [1, 10]$ definita come segue:

$$g(x) = \frac{1}{2}\left(27\log x - x + 3\right).$$

Poiché $g(1) = 1$, $g(10) = 10$ e $g^{'}(x) = \frac{1}{2}\left(\frac{27\log e}{x} - 1\right) > 0$ per $x \in [1, 10]$ (ossia g è crescente), g è un omeomorfismo di $[1, 10]$ e lo è anche la sua inversa $h := g^{-1}$ (che sarà invece decrescente in $(1, 10)$). Inoltre, $g(x) > x$ per $x \in [1, 10]$, questo perchè $s(x) = g(x) - x$ è tale che $s(1) = s(10) = 0$ e $s^{'}(x) \leq 0$ per $x \in [1, 9\log e]$, dunque $s(x) > 0$ per $x \in [1, 10]$. Da ciò segue che $h(x) < x$ per $x \in (1, 10)$.

Sia ora $T = U(1, 10)$ e X_T la variabile aleatoria discreta avente due possibili esiti, ovvero T e $h(T)$, ottenuti rispettivamente con probabilità $\frac{1}{3}$ e $\frac{2}{3}$.

Osserviamo che:

- nessuna delle variabili aleatorie X_T è Benford, questo perché, per ogni T, vi è un atomo di massa 2/3 e perciò $\sup_{t\in[1,10)} \left|F_{S(X_T)}(t) - \log t\right| \geq \frac{1}{3}$;
- per ogni $t \in [1, 10)$

$$\begin{aligned}
\mathbb{P}\left(\mathbb{E}X_T \leq t\right) &= \mathbb{P}(X_T = T)\mathbb{P}\left(T \leq t\right) + \mathbb{P}\left(X_T = h\left(T\right)\right)\mathbb{P}\left(h\left(T\right) \leq t\right) \\
&= \frac{1}{3}\cdot\mathbb{P}\left(T \leq t\right) + \frac{2}{3}\cdot\mathbb{P}\left(h\left(T\right) \leq t\right) \\
&= \frac{1}{27}(t-1) + \frac{2}{3}\cdot\mathbb{P}\left(T \leq g(t)\right) \\
&\overset{g(t)\in[1,10)}{=} \frac{1}{27}(t-1) + \frac{2}{3}\cdot\frac{1}{9}\left(g\left(t\right) - 1\right) \\
&= \frac{1}{27}(t-1) + \frac{2}{27}\left(g\left(t\right) - 1\right) \\
&= \log t,
\end{aligned}$$

ovvero $\mathbb{E}X_T$ è Benford.

Il prossimo lemma mostra che la proporzione limite di volte in cui una successione di m-campioni P-casuali cade in un boreliano B è, con probabilità uno, il valore medio rispetto alla misura di probabilità $\mathbb{P}$ dell'insieme B, cioè è uguale a $\mathbb{E}P(B)$. Si noti che ciò non è semplicemente un corollario diretto della Legge Forte dei Grandi Numeri classica, poiché le variabili casuali nella successione non sono necessariamente indipendenti.

Lemma 2.2. *Sia P una r.p.m. e $(X_1, X_2, \dots)$ una successione di m-campioni P-casuali per qualche $m \in \mathbb{N}$. Allora $\forall B \in \mathcal{B}^+$*

$$\frac{\#\{1 \leq n \leq N : X_n \in B\}}{N} \xrightarrow{q.c.} \mathbb{E}P(B) \quad \textit{per } N \to \infty.$$

Dimostrazione. Fissiamo $B \in \mathcal{B}^+$, $j \in \mathbb{N}$ e poniamo $Y_j = \#\left\{1 \leq i \leq m : X_{(j-1)m+i} \in B\right\}$. E' chiaro che

$$\lim_{N\to\infty} \frac{\#\{1 \leq n \leq N : X_n \in B\}}{N} = \frac{1}{m} \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Y_j \tag{2.5}$$

se il limite a destra esiste. Osserviamo che $Y_j|P_j$ conta quante tra le $X_{(j-1)m+1}, \dots, X_{jm}$ cadono nell'insieme B, quindi può essere pensata come una binomiale di parametri m (stiamo considerando m-campioni P-casuali) e $P_j(B)$ (per la Definzione 2.2 (i)), dunque $\mathbb{E}[Y_j|P_j] = mP_j(B)$. A questo punto ci interessa calcolare $\mathbb{E}[Y_j]$ che è data da

$$\mathbb{E}[Y_j] = \mathbb{E}\left[\mathbb{E}\left[Y_j|P_j\right]\right] = \mathbb{E}[mP_j(B)] = m\mathbb{E}[P_j(B)] = m\mathbb{E}P_j(B),$$

e poiché le P_j hanno la stessa distribuzione di P si ha che

$$\mathbb{E}[Y_j] = m\mathbb{E}P(B). \tag{2.6}$$

Per la Definzione 2.2 (ii), le Y_j sono indipendenti; sono inoltre uniformemente limitate, poiché $0 \leq Y_j \leq m\ \forall j$, dunque $\mathbb{E}[Y_j^2] \leq m^2$. Ma allora $\sum_{j=1}^{\infty} \mathbb{E}[Y_j^2]/j^2 < \infty$.
Per di più, per (2.6) tutte le Y_j hanno stessa media $m\mathbb{E}P(B)$. Dunque, per il Corollario[5, 5.1],

$$\frac{1}{n} \sum_{j=1}^{n} Y_j \xrightarrow{q.c.} m\mathbb{E}P(B) \quad \text{per } n \to \infty \tag{2.7}$$

e la conclusione segue da (2.5) e (2.7).

□

Ora siamo pronti per enunciare una legge limite statistica (Teorema 2.19 qui sotto), che è il teorema del tipo "Limite Centrale" per le cifre significative menzionato sopra. In termini generali, questa legge afferma che se le distribuzioni di probabilità vengono selezionate casualmente e successivamente vengono presi campioni casuali da ciascuna di queste distribuzioni in modo tale che l'intero processo sia neutrale rispetto alla scala o alla base, allora le frequenze delle cifre significative del campione combinato convergeranno alla distribuzione logaritmica. Questo teorema può aiutare a spiegare e prevedere l'apparizione della legge di Benford in miscele di dati tabulati, come i dati combinati delle tabelle di Benford (così come il suo dataset individuale di numeri raccolti dai giornali).
Per trarre qualsiasi conclusione riguardo la legge di Benford per il processo di campionamento da distribuzioni differenti, chiaramente deve esserci qualche restrizione sulla r.p.m. sottostante che genera la procedura di campionamento. Altrimenti, se la r.p.m. è, per esempio, $U(0,1)$ con probabilità uno, allora qualsiasi sequenza risultante di m-campioni P-casuali sarà i.i.d. $U(0,1)$, e quindi quasi sicuramente non sarà Benford. Una supposizione naturale da fare riguardo una r.p.m. in questo contesto è che in media la r.p.m. abbia cifre significative indipendenti dalla scala o dalla base. Ricordiamo che $\mathcal{S}$ denota la σ-algebra delle cifre significative (Definizione 1.3).

Definizione 2.3. *Sia $\mathcal{A} \supset \mathcal{S}$ una σ-algebra su $\mathbb{R}^+$. Una misura di probabilità P su $(\mathbb{R}^+, \mathcal{A})$ ha cifre significative invarianti per cambiamenti di scala se*

$$P(aA) = P(A) \quad \forall a > 0 \ , \forall A \in \mathcal{S},$$

o, equivalentemente, se $\forall m \in \mathbb{N}$, $\forall d_1 \in \{1,\dots,9\}$, $\forall d_j \in \{1,\dots,9\}$, $j \geq 2$ e $\forall a > 0$,

$$P\left(\{x : D_j(ax) = d_j \ per\ j = 1,\dots,m\}\right) = P\left(\{x : D_j(x) = d_j \ per\ j = 1,\dots,m\}\right).$$

Una misura di probabilità P su $(\mathbb{R}^+, \mathcal{A})$ ha cifre significative invarianti per cambiamenti di base se

$$P(A^{1/n}) = P(A) \quad \forall n \in \mathbb{N} \ , \forall A \in \mathcal{S}.$$

Definizione 2.4. *Una r.p.m. P possiede cifre (decimali) significative indipendenti dalla scala se*

$$\mathbb{E}P(aA) = \mathbb{E}P(A) \quad \forall a > 0 \ , \forall A \in \mathcal{S},$$

o, equivalentemente, se la misura di probabilità di Borel $\mathbb{E}P$ ha cifre significative invarianti per cambiamenti di scala.

Analogamente, una r.p.m. P possiede cifre (decimali) significative indipendenti dalla base se

$$\mathbb{E}P(A^{1/n}) = \mathbb{E}P(A) \quad \forall n \in \mathbb{N} \ , \forall A \in \mathcal{S},$$

ovvero se $\mathbb{E}P$ ha cifre significative invarianti per cambiamenti di base.

Ci servono ora alcuni risultati per arrivare alla dimostrazione del Teorema 2.19. Il Teorema 2.17 ci dà una caratterizzazione delle misure di probabilità aventi cifre significative invarianti per cambiamenti di scala, mentre il Teorema 2.18 fornisce una caratterizzazione delle misure di probabilità aventi cifre significative invarianti per cambiamenti di base.

Teorema 2.17. *Una misura di probabilità P su $(\mathbb{R}^+, \mathcal{A})$ con $\mathcal{A} \supset \mathcal{S}$ ha cifre significative invarianti per cambiamenti di scala se e solo se $P(A) = \mathbb{B}(A)$ per ogni $A \in \mathcal{S}$, ossia se e solo se P è Benford.*

Dimostrazione. Si scelga una misura di probabilità P su $(\mathbb{R}^+, \mathcal{A})$ e si denoti con P_0 la sua restizione a $(\mathbb{R}^+, \mathcal{S})$. Poniamo $Q = P_0 \circ \ell^{-1}$ con ℓ definita come nel Lemma 1.2. Per il Lemma 1.2, Q è una misura di probabilità su $([0,1), \mathcal{B}(0,1])$. Inoltre, per la corrispondenza stabilita da ℓ, l'affermazione

$$P_0(aA) = P_0(A) \quad \text{per ogni } a > 0, A \in \mathcal{S} \tag{2.8}$$

è equivalente a

$$Q(\langle B + x \rangle) = Q(B) \quad \text{per ogni } x \in \mathbb{R}, B \in \mathcal{B}[0,1), \tag{2.9}$$

dove $\langle B + x \rangle = \{\langle b + x \rangle : b \in B\}$. Sia X una variabile aleatoria tale che la distribuzione di $\langle X \rangle$ sia Q. Allora (2.9) afferma che, per ogni $x \in \mathbb{R}$, le distribuzioni di $\langle X \rangle$ e $\langle X + b \rangle$ coincidono. Per il Teorema 1.8 (ii) (si osservi che una variabile aleatoria costante è indipendente da ogni altra variabile aleatoria), questo avviene se e solo se X è u.d. mod 1, ossia $Q = \lambda_{0,1}$.
Dunque (2.8) è equivalente a $P_0 \circ \ell^{-1} = \lambda_{0,1} = \mathbb{B} \circ \ell^{-1}$. Per il Lemma 1.2, (2.8) è equivalente a $P_0 = \mathbb{B}$.

□

Teorema 2.18. *Una misura di probabilità P su $(\mathbb{R}^+, \mathcal{A})$ con $\mathcal{A} \supset \mathcal{S}$ ha cifre significative invarianti per cambiamenti di base se e solo se, per un certo $q \in [0,1]$,*

$$P(A) = q\delta_1(A) + (1-q)\mathbb{B}(A) \quad \textit{per ogni } A \in \mathcal{S}. \tag{2.10}$$

Dal Teorema 2.17 e dal Teorema 2.18 segue immediatamente la seguente proposizione.

Proposizione 2.2. *Per ogni r.p.m. P le seguenti affermazioni sono equivalenti:*

i. P *ha cifre significative indipendenti dalla scala;*

ii. $P \circ S^{-1}(\{1\}) = 0$ *e* P *ha cifre significative indipendenti dalla base;*

iii. $\mathbb{E}P(A) = \mathbb{B}(A)$ *per ogni* $A \in \mathcal{S}$, *ossia* $\mathbb{E}P$ *è Benford.*

Le r.p.m. con cifre significative indipendenti dalla scala o dalla base sono facili da costruire matematicamente, tuttavia, negli esempi della vita quotidiana, non dovrebbero essere date per scontate. Ad esempio, supponiamo di selezionare casualmente aziende produttrici di bevande in Europa e osservare quanti volumi metrici di m prodotti produce ogni azienda. Questo processo verosimilmente non produrrà dati con cifre significative indipendenti dalla scala, poiché i volumi sono strettamente legati all'unità di misura utilizzata per misurarli, ovvero i litri. La conversione dei dati in un'altra unità, come i galloni, probabilmente produrrebbe una distribuzione delle frequenze delle cifre significative radicalmente diversa.
Come verrà mostrato nel prossimo teorema, l'indipendenza dalla scala o dalla base di una r.p.m. implica che le sequenze di m-campioni P-casuali sono Benford quasi certamente. Un punto cruciale nella definizione di una r.p.m. P con cifre significative indipendenti dalla scala o dalla base è che le singole realizzazioni di P non sono tenute ad avere cifre significative invarianti rispetto alla scala o alla base. Infatti, spesso accade che quasi certamente nessuna delle probabilità casuali abbia cifre significative invarianti rispetto alla scala o alla base; è solo in media che il processo di campionamento non favorisce una scala o una base rispetto a un'altra.
Si ricordi che $P \circ S^{-1}(\{1\}) = 0$ è l'evento $\{\omega \in \Omega : P(\omega)(S = 1) = 0\}$.

Teorema 2.19. *Sia P una r.p.m. Si assuma che essa abbia cifre significative indipendenti dalla scala o che abbia cifre significative indipendenti dalla base e $P \circ S^{-1}(\{1\}) = 0$ con probabilità 1. Allora, per ogni $m \in \mathbb{N}$, ogni successione $(X_1, X_2, \dots$ di m-campioni P-casuali è Benford con probabilità 1 se per ogni $t \in [1, 10)$*

$$\frac{\#\{1 \le n \le N : S(X_n) \le t\}}{N} \xrightarrow{q.c.} \log t \quad \textit{per } N \to \infty.$$

Dimostrazione. Si inizi assumendo che P abbia cifre significative indipendenti dalla scala, cioè $\mathbb{E}P$ ha cifre significative invarianti per cambiamenti di scala. Per il Teorema 2.17, $\mathbb{E}P$ è Benford.
Il Lemma 2.2 implica che per ogni successione $(X_1, X_2, \dots)$ di m-campioni

P-casuali e ogni $t \in [1, 10)$

$$\frac{\#\{1 \leq n \leq N : S(X_n) \leq t\}}{N} = \frac{\#\left\{1 \leq n \leq N : X_n \in \bigcup_{k\in\mathbb{Z}} 10^k[1,t]\right\}}{N}$$
$$\xrightarrow{q.c.} \mathbb{E}P\left(\bigcup_{k\in\mathbb{Z}} 10^k[1,t]\right) = \log t \quad \text{per } N \to \infty.$$

Si assuma ora che P abbia cifre significative indipendenti dalla base e $P \circ S^{-1}(\{1\}) = 0$ con probabilità 1. Allora

$$\mathbb{E}P\left(S^{-1}(\{1\})\right) = \int_\Omega P \circ S^{-1}(\omega)(\{1\})\, d\mathbb{P}(\omega) = 0,$$

dunque, rimpiazzando P con $\mathbb{E}P$ in (2.10) vediamo che $q = 0$. Ciò prova che $\mathbb{E}P$ è Benford e il resto dell'argomentazione rimane uguale.

□

Un punto principale del Teorema 2.19 è che esistono molte procedure di campionamento naturali che portano alla stessa distribuzione logaritmica. Questo potrebbe aiutare a spiegare come i diversi risultati osservati da Newcomb, Benford, Knuth e Nigrini abbiano portato tutti alla stessa legge. Potrebbe anche spiegare perché i campioni di dati raccolti dalle prime pagine dei giornali [1] e o dai dati contabili estesi [12] tendano spesso a seguire la legge di Benford, poiché in ciascuno di questi casi, varie distribuzioni vengono campionate in modo presumibilmente casuale. In un giornale, ad esempio, il primo articolo potrebbe contenere statistiche sulla crescita della popolazione, il secondo sui prezzi delle azioni, il terzo sulla superficie forestale.
La giustificazione dell'ipotesi di indipendenza dalla scala o dalla base delle cifre significative nella pratica è analoga alla giustificazione dell'ipotesi di indipendenza e identica distribuzione di quando si applicano la Legge Forte dei Grandi Numeri o il Teorema del Limite Centrale ai processi reali: nessuna delle due ipotesi può essere formalmente provata, ma sembrano essere ipotesi ragionevoli per molte procedure di campionamento della vita reale.
Molte costruzioni standard di una r.p.m. hanno automaticamente cifre significative indipendenti dalla scala e dalla base, e quindi soddisfano la legge di Benford nel senso del Teorema 2.19.

Esempio 2.11
Si richiami la costruzione di Dubins-Freedman già descritta nell'Esempio 2.8. Segue dal Teorema 1.5 e dal Lemma[6, 9.28] che $\mathbb{E}P$ è Benford, quindi, per la Proposizione 2.2, P ha cifre significative indipendenti dalla scala e dalla

base. Si noti, tuttavia, che con probabilità uno, P non avrà cifre significative invarianti rispetto alla scala o alla base. Ciò accade solo in media, ma il Teorema 2.19 ci garantisce che ciò è sufficiente affinché il campionamento P-casuale generi successioni quasi certamente Benford.
Nella costruzione di Dubins-Freedman, il fatto che $F_P(10^{1/2})$, $F_P(10^{1/4})$, $F_P(10^{3/4})$, ecc., siano scelti uniformemente dai rispettivi intervalli non è cruciale: se Q è una qualunque misura di probabilità su $(0,1)$, e i valori di $F_P(10^{1/2})$, ecc., sono scelti indipendentemente secondo una versione opportunamente scalata di Q allora, a condizione che Q sia simmetrica rispetto a $1/2$, per l'r.p.m. così generata, $\mathbb{E}P$ sarà ancora di Benford; si veda [6, Teor. 9.29].
I processi nel mondo reale spesso mostrano naturalmente questa simmetria: molti dati possono essere registrati indifferentemente usando certe unità o i loro reciproci, ad esempio, in miglia per gallone contro galloni per miglio, o nelle "candele per watt" contro i "watt per candela"(Benford). Ciò suggerisce che la distribuzione di $\log S$ dovrebbe essere simmetrica rispetto a $1/2$.

Infine, è importante sottolineare che molte r.p.m. e processi di campionamento non seguono la legge di Benford e, di conseguenza, sono necessariamente influenzati dalla scala e dalla base.

Conclusione

La presente tesi ha esplorato le caratteristiche della legge di Benford in relazione a variabili aleatorie e processi stocastici. Sono stati volutamente tralasciati i sistemi deterministici unidimensionali (che prendono la forma di equazioni alle differenze per sistemi a tempo discreto e di equazioni differenziali unidimensionali per sistemi a tempo continuo) che generano la legge di Benford.
Nel primo capitolo abbiamo costruito le fondamenta teoriche per analizzare le relazioni tra la legge di Benford e le distribuzioni di probabilità. In particolare, abbiamo introdotto concetti come la funzione dei significandi e la σ-algebra dei significandi, strumenti essenziali per formalizzare il comportamento delle cifre significative in termini probabilistici. Questa parte iniziale della tesi ha anche chiarito che, sebbene nessuna delle distribuzioni classiche come la normale, l'esponenziale o la uniforme segua esattamente la legge di Benford, alcune di esse possono avvicinarsi a tale distribuzione in condizioni particolari, come evidenziato dai teoremi riguardanti la distribuzione di Pareto. Abbiamo inizialmente rivolto la nostra attenzione a formalizzare la legge di Benford su $(\mathbb{R}^+, \mathcal{S})$, per poi spostare il nostro interesse su $([0,1), \mathcal{B}[0,1))$ e studiare il legame tra distribuzione uniforme modulo uno e legge di Benford.
Il secondo capitolo ha invece esplorato come la legge di Benford emerga come distribuzione limite in vari contesti, ponendo l'attenzione sull'analisi delle potenze, dei prodotti e delle somme di variabili aleatorie. Un risultato cruciale, qui dimostrato, è che determinate operazioni su variabili aleatorie indipendenti e identicamente distribuite (i.i.d.), come prodotti, potenze o combinazioni lineari, portino ad una convergenza alla distribuzione di Benford. Infine, abbiamo enunciato e dimostrato un teorema simile al Teorema del Limite Centrale: esso afferma che se le distribuzioni di probabilità vengono selezionate casualmente e successivamente vengono presi campioni casuali da ciascuna di queste distribuzioni, in modo tale che l'intero processo sia neutrale rispetto alla scala o alla base, allora le frequenze delle cifre significative del campione combinato convergeranno alla legge di Benford. Tale teorema ci permette di spiegare perché anche se alcuni dataset non seguono la legge

di Benford, combinazioni di essi lo fanno.

Bibliografia

[1] F. Benford, *The law of anomalous numbers*, Proc. Amer. Philosophical Soc., 78(4):551–572, 1938.

[2] A. Berger e T. P. Hill, *An Introduction to Benford's Law*, Princeton University Press, 2015.

[3] A. Berger e T. P. Hill, *The Mathematics of Benford's Law - A Primer*, Statistical Methods and Applications, 2020.

[4] L. Berton, *He's got their number: Scholar uses math to foil financial fraud*, The Wall Street Journal, July 10, 1995.

[5] Y. S. Chow and H. Teicher, *Probability Theory. Independence, Interchangeability, Martingales*, Springer Texts in Statistics, Springer, New York, terza edizione, 1997.

[6] L. E. Dubins and D. A. Freedman, *Random distribution functions*, Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Parte 1, pp. 183–214. University of California Press, Berkeley, 1967.

[7] D. E. Knuth, *The Art of Computer Programming*, Addison-Wesley, Reading, Mass.-London-Amsterdam, seconda edizione, 1975.

[8] M. T. Lacey e W. Philipp, *A note on the almost sure central limit theorem*, Statist. Probab. Lett., 9(3):201–205, 1990.

[9] S. Newcomb, *Note on the Frequency of Use of the Different Digits in Natural Numbers*, American Journal of Mathematics, 4(1-4):39–40, 1881.

[10] M. J. Nigrini, *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*, John Wiley & Sons, Hoboken, NJ, 2012.

[11] M. J. Nigrini, *A taxpayer compliance application of Benford's law*, The Journal of the American Taxation Association, 18(1):72–91, 1996.

[12] M. J. Nigrini, *The Detection of Income Tax Evasion through an Analysis of Digital Distributions*, PhD dissertation, Department of Accounting, University of Cincinnati, Cincinnati, OH, 1992.

[13] R. A. Raimi, *The first digit problem*, Amer. Math. Monthly, 83(7):521–538, 1976.

[14] M. Sambridge, H. Tkalčić e A. Jackson, *Benford's law in the natural sciences*, Geophysical Research Letters, 2010.