國立政治大學理學院應用物理研究所

碩士論文

Graduate Institute of Applied Physics, College of Science

National Chengchi University

Master Thesis

二位元序列中連續 1 的記憶效應初探

A primitive study of memory effect of run of ones in binary sequences

李雨純

Yu-Chun Li

指導教授：馬文忠博士

Advisor：Wen-Jong Ma

中華民國一一〇年一月

January, 2021

DOI:10.6814/NCCU202100266

誌謝

在這研究所的這段時間中，總是受到許多人的幫忙，讓我在有興趣卻沒有深厚基礎的學科中，得以探索與汲取學識，並且順利完成這個求學階段。

首先，非常感謝我的指導教授馬文忠老師，不只在實驗階段遇到困難時，很有耐心的給予方向指引，更在數學上給予很多啟發，讓我對整個實驗模擬能有更嚴謹的定義與想法，也感謝老師在我撰寫論文期間，常給予遇到困難的我許多的指導與修正，讓我得以順利完成這份論文。再者感謝所長與所辦小姐，總是溫暖的關心所上的每一位學生，也感謝所上的老師，在我修課期間，總能在課堂中給予許多專業知識與思考，奠定成長的基礎；也感謝在論文口試時，給予寶貴意見與指正的蔡尚岳老師與國立東華大學陳企寧老師，讓這份論文能更完整的呈現。

另外也感謝所上的同學、學長姊與學弟妹，在大家的陪伴與幫忙下能順利完成這學業，也感謝王世傑學長、江林宗叡學弟、陳政國學弟、吳培煜同學，給予我在實驗上與論文撰寫時的建議與幫助，也感謝我的同事李柏翰老師，給予我口試時的意見與指導。

最後感謝母校-政治大學與所上-應用物理研究所，提供一個讓我成長的環境與人文，讓我在此階段能夠成長並收穫良多。

DOI:10.6814/NCCU202100266

Abstract

In a complex financial system, there are many changing factors that are responsible for the time evolution of the market. There have been efforts to reveal those factors, which result in the observations of many “stylized facts”. One important observation is the presence of power laws in the distributions of various quantities. In this thesis, the author chose to explore such properties on a fundamental level to find whether, or how power laws can be generated in simple Brownian motion. The author uses a one-dimensional model to explore the fall and rise of stock prices, treating them as 0 and 1 in binary, respectively. The time evolution of the price changes of a stock is then realized as a binary sequence. The analysis goes to find the distributions of runs of ones (sections of consecutive ones) in binary sequences. In our model, the next bit (price change) at each time step is determined, either at random (trading without memory) or in accord with the history (trading with memory). The time sequence eventually converges to some steady state. It is found in this study that, for the sequences with memory, the mean arrival time of convergence is a power law function of the memory length. After scale transformation, the curves of the probability density distributions of arrival times for different memory lengths overlap with each other nicely. The result suggests the power-law properties in the distributions of financial data may be related to some underlying scaling behavior of the system.

DOI:10.6814/NCCU202100266

摘 要

在複雜的金融系統中，許多變動因素影響著市場，人們在揭開這些因素的努力中觀察到許多典型化的實況(stylized facts)；其中一個重要的結果為:各種計量的分布呈現著冪次定律(power-law)。本論文回歸到一個基本的問題:在簡單的布朗運動中是否能夠、或者如何產生冪次定律的特性；藉由將股價的下跌與上漲視為二位元的 0 與 1，將一支股票的時間演變，視為一個二位元序列，所要分析的就變成序列中連續 1 的分布性質。我們的模型中，在某一時間點的下一個位元(股價變動)，可以是任意產生(交易不依靠記憶)或是根據交易歷史來決定(根據記憶進行交易)；這樣的時間序列最終會收斂到某種穩定狀態。我們發現，有記憶的序列，其抵達收斂狀態的平均時間與記憶長度呈現冪次定律的關係，而在不同記憶長度下得到的抵達時間之機率密度函數，在經過尺度轉換後，其曲線彼此間有很好的重疊性。此結果顯示金融數據所呈現的冪次定律分布，或可連結到系統內在的尺度不變性。

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100266

目錄

誌謝 ……I

Abstract ……II

摘要 ……III

第一章 緒論 ……1

第二章 理論背景與原理 ……4

2.1 與事件次數發生相關的分布 ……4

2.1.1 伯努利試驗與二項分布 (Binomial distribution) ……4

2.1.2 卜瓦松分布(Poisson distribution) ……5

2.1.3 幾何分佈(Geometric distribution) ……6

2.1.4 負二項分布 (negative binomial) ……7

2.2 與時間有關的連續型分布 ……7

2.2.1 指數分佈(Exponential distribution) ……7

2.2.2 布朗運動與一維隨機行走(Brownian Motion) ……8

2.2.3 常態分佈(高斯分佈)(Normal distribution) ……9

2.2.4 萊維穩定分布 (Lévy Stable Distribution) ……10

第三章 二位元序列中的連續列探討 ……11

3.1 出現給定長度連續 1 所需的等待時間 ……11

3.1.1 第一次出現給定長度連續 1 所需的等待時間 ……12

3.1.2 第 r 次出現給定長度連續 1 所需的等待時間 ……13

3.2 樣本空間的分析與模擬 ……14

3.2.1 完整連續列的樣本空間 ……15

3.2.2 次連續列 (subsequence)的樣本空間 ……15

3.2.3 電腦模擬:以單一的長序列產生樣本空間中短序列連續 1(run of ones)的分布 ……17

DOI:10.6814/NCCU202100266

第四章 二位元序列的隨機過程與記憶效應……19
4.1 固定記憶長度 m 的隨機過程，連續列的分布過程為指數函數 ……20
4.2 全記憶的模擬:立即更新與變動更新 ……21
4.3 有限記憶($m > 1$ 且固定)二位元序列的長時間收斂分析 ……25
4.3.1 有限記憶二位元序列的時間演進 ……25
4.3.2 固定記憶二位元序列的隨機過程分析：轉移矩陣(transition matrix)……29
4.3.3 固定記憶二位元序列的隨機過程分析：$T_{arrival}$ ……31
第五章 結論……34
附錄 A ……35
附錄 B ……40
附錄 C ……44
附錄 D ……45
附錄 E ……46
參考文獻 ……48

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100266

圖表目錄

圖 1.1 ……2
圖 1.2 ……3
圖 3.1 ……11
表 3.1……16
圖 3.2 ……17
圖 4.1……20
圖 4.2……21
圖 4.3……21
圖 4.4……22
圖 4.5……23
圖 4.6……24
圖 4.7……26
圖 4.8……26
圖 4.9……27
圖 4.10 ……27
圖 4.11 ……28
圖 4.12 ……28
圖 4.13 ……32
圖 4.14 ……32
圖 4.15 ……33

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100266

第一章 緒論

近數十年來，由於網路的連結與計算工具的發展，使得財務金融的運作有了巨大的變化。在社會科學與自然科學跨領域的努力下，對複雜金融市場的現象有更深入的理解。

早在 1900 年，法國數學家 Bachelier 就已由股票價格的變動提出隨機行走(random walk)的概念[1]，同時期(1905 年)愛因斯坦則由布朗運動的描述提出相同的概念。其後將物理學中的布朗運動描述微粒子的動態軌跡模型，與股市分析做連結，逐步成為一個跨領域的議題。金融系統的研究在 20 世紀後期因為許多物理學家的加入，結合了社會科學與自然科學的跨領域研究，形成金融物理學(econophysics)此一新興學門[2]。

1959 年物理學家 Osborne[3]在他的研究中提出股票價差(或收益 return)的分布並不是呈現常態分布。學者 Fama（1965）研究價格及報酬的分布型態時，發現尖峰厚尾的分布[4]。Mantegna 和 Stanley(1995) [5]以及 Bouchaud 和 Potters(1997) [2][6]的研究進一步確認金融市場中各商品所得到的報酬率分布都是會呈現厚尾分布。其後 Stanley 等人(1999) [7]以及 Bouchaud 和 Potters 等人(1999)[8]研究股票市場的交互相關性，以整體市場為著眼，將個股間股價波動的交互影響予以量化。2004 年 Ma, Hu 和 Amritkar 針對股票交互相關性建立耦合隨機行走(coupled random walk)的模型[9]。Ma, Wang, Chen 和 Hu 又在 2013 利用股價收益與布朗運動位移的類比，計算出市場的動力參數[10]。近期有關極端事件複雜系統的分析開始引起學界的興趣，例如 Kishore, Santhanam 和 Amritkar 在 2012 年的文章在於網路極端事件分析，討論極端事件的

DOI:10.6814/NCCU202100266

性質[11]。Chen 和 Ma[12]施奕甫[13]與吳培煜[14]分析股票群體報酬在不同時間尺度下的特性，施[13]與吳[14]比較台灣與美國股市的數據，討論事件的持續時間與跨天效應的影響。

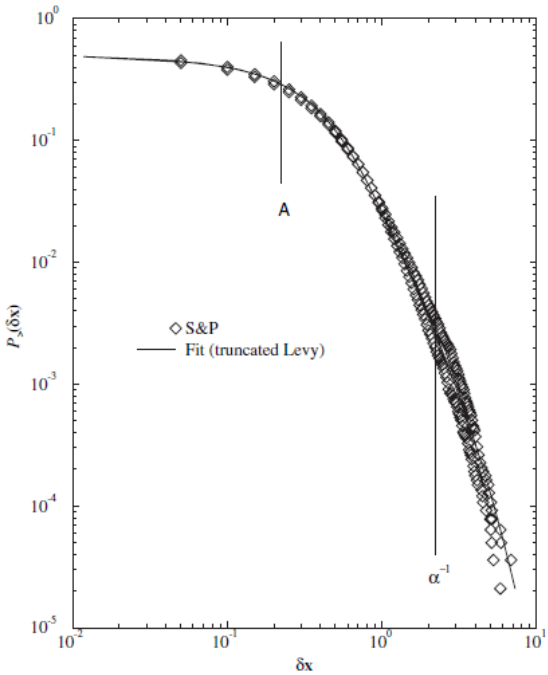


圖 1.1：S&P500 指數 15 分鐘的價差分布，具有厚尾性質。(本圖摘自參考文獻[2] Fig.5.11.,原圖來自 J.-P. Bounchaud, ,M. Potters 所著 *Théorie des risques financiers* 一書)

厚尾分布為金融數據的一項典型化的實況(stylized facts)[2][6]，具有冪次定律(power-law)的特徵 (見圖 1.1)。冪次定律也出現在股價上升(或下跌)事件的持續時間分布 [13] [14] (見圖 1.2)；此冪次分布與收益的厚尾是否有關，為一項值得探討的問題。而一個隨機產生的序列的持續時間是否可能呈現冪次分布，便是本研究中所關心的議題。本論文試著將影響著股票市場的複雜系統，回歸到簡單的布朗運動，以一維隨機模型簡化為探討股價的下跌與上漲兩種狀況，分別視為二位元中的 0 與 1，藉以尋求可能產生冪次定律的情況。

DOI:10.6814/NCCU202100266

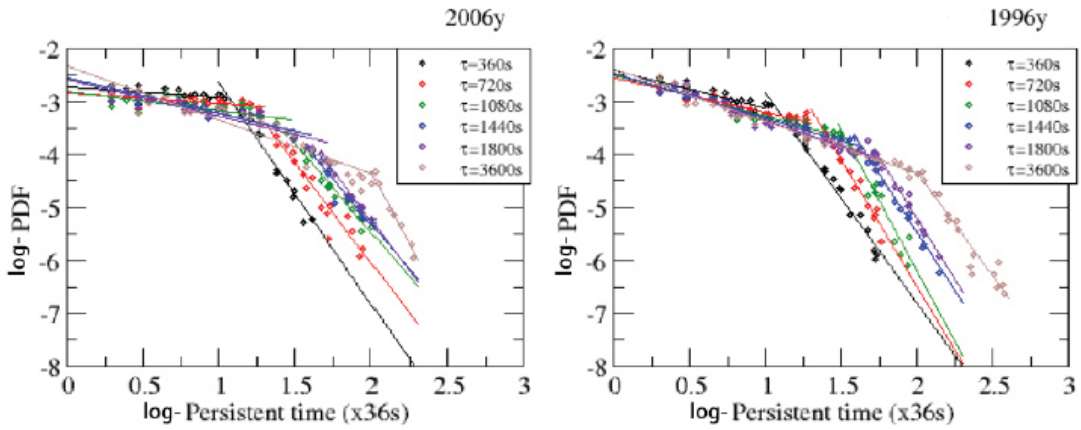


圖 1.2：台灣市場(TWII)2006 年(左圖)與美國市場(SP500)1996 年(右圖)分別各取 386 支股票與 345 支股票的數據，計算上漲事件的持續時間得到的分布機率密度函數(probability density function, PDF)對數圖。計算報酬所採用的時間間隔(τ)的大小不論如何改變，皆會得到冪次分布，並在 τ 處發生轉折的情形(本圖摘自吳[14]的碩士論文 Figure 4.2)。

本文以此為出發探討，將二位元序列中，1 的連續列(run of ones)段落比擬為連續上漲的區間，觀察其機率分布。此序列相關的組合計算分析於近期文獻中有詳細的探討[15][16][17]。我們的模型產生隨機序列的方式，又可分為無記憶性質的過程，與有記憶性質下的視窗移動(scan)，就此兩種不同狀態下，藉由電腦模擬進行分析，在有記憶的 0 與 1 序列中，下一個位元的數值是由前一個記憶視窗中的 0 與 1 分布而決定，此序列會收斂到此視窗中為全為 0 或全為 1 的狀態。我們發現，連續 1 的字串長度分布基本上為指數函數，而進一步分析在有記憶性質下的視窗移動模擬中，各個字串長度的平均收斂時間，隨視窗大小成冪次形式遞增，且經由尺度變換後，分布機率密度函數曲線重疊為一。

本論文在第二章中定義基本的機率分布型態，在第三章中由近期文獻有關二位元序列的組合計算出發，整理序列中連續 1 相關的統計量。第四章則為時間序列的電腦模擬分析，最後一章為討論與結論。

DOI:10.6814/NCCU202100266

第二章 理論背景與原理

金融數據的分布常呈現厚尾、冪次分布(power-law)，這些分布特性的背後來源為一項值得探討的問題。一個隨機產生的二位元序列，其持續時間是否可能呈現冪次分布，便是本研究中所關心的議題。有關二位元序列中，連續 1 的組合計算分析於近期文獻中有詳細的探討[15][16][17]，於這些分析中有關的組合分布，我們於本章做一個完整的整理，內容包含二項分布(第 2.1 節)、卜瓦松分布(第 2.2 節)、指數分布(第 2.3 節)、幾何分布(第 2.5 節)、負二項分布(第 2.6 節)等。

描述事件的發生(二位元序列中 1 的出現)的各種情況，第一個情況為獨立事件且和時間無關，其事件發生(1 的出現)次數的機率分布為二項分布；若考慮第一次事件發生(1 的出現)所需經過的試驗次數(0 的出現次數、再加 1)的機率分布為幾何分布 ；若考慮的是第 r 次事件發生所需經過的試驗次數(0 的出現次數、再加 r)的機率分布為負二項分布。第二個情況為獨立事件發生的機率與離前一個事件發生的時距成正比，其事件發生(1 的出現)次數的機率分布為卜瓦松分布；若描述第一次事件發生的時間的機率分布，則為指數分布[18]。其中時間為一個連續的量，而用以描述時間的量，除了指數分布以外，在金融數據中，常被提及的常態分布與萊維分布亦整理於本章中。

2.1 與事件次數發生相關的分布

2.1.1 伯努利試驗與二項分布(Binomial distribution)

某隨機現象可以重複，每觀察一次這種隨機現象稱為進行一次隨機試驗，只有兩種可能結果的隨機試驗稱為伯努力試驗(Bernoulli trial)。

DOI:10.6814/NCCU202100266

n 個獨立的伯努利試驗中，成功機率皆為 p ，失敗機率皆為 $q=1-p$ ，隨機變數 X 為總共成功的次數，以 $X \sim B(n,p)$ 表示。二項分布為離散型的分布，其統計性質如下：

(1)機率質量函數為 $f\left(x\right)=C_x^n p^x\left(1-p\right)^{n-x}$, $x=0,1,\cdots,n$

(2)期望值為 $E(X)=np$

(3)變異數為 $Var(X)=np(1-p)$

2.1.2 卜瓦松分布(Poisson distribution)

在特定時間區間內，觀察事件發生的次數，若具有以下性質，則稱為卜瓦松過程(Poisson process)：

(1) 在一個短時間區間 Δt，發生一次事件的機率與 Δt 成正比。

(2) 在短時間內發生兩次以上的機率可以忽略。

(3) 在不重疊的時間段落裡，事件各自發生的次數是獨立的。

設隨機變數 X_n 有 $B(n,p_n)$ 分佈，$n \geq 1$，且滿足 $\lim_{n\to\infty} np_n=\eta$ ，$0<\eta<\infty$ ，則

$$\lim_{n\to\infty} P(X_n=k)=\frac{e^{-\eta}\eta^k}{k!}, k=0,1,\cdots \tag{2.1}$$

n 很大時，通常以卜瓦松分佈之機率值為二項分佈之機率值的近似值(稱為卜瓦松近似)。$np \leq 7$ 時，卜瓦松分佈對二項分佈提供一個很好的近似值。但 np 過大時，不宜用卜瓦松分佈來當二項分佈的近似值。卜瓦松分布為離散型的分布，其統計性質如下：

(1) 機率質量函數為 $f\left(x\right)=\dfrac{e^{-\eta}\eta^x}{x!}$, $x=0,1,\cdots$

DOI:10.6814/NCCU202100266

(2) 期望值為 $E(X)=\eta$

(3) 變異數為 $Var(X)=\eta$

2.1.3 幾何分佈(Geometric distribution)

持續投擲一枚正面機率 $p>0$ 的銅板，每次投擲皆為獨立，直到出現第一次正面才停止，其總投擲次數為 X，稱為有參數 p 之幾何分布，以 $X\sim Geo(p)$ 表示，幾何分布為離散型的分布，其統計性質如下：

(1) 機率質量函數為 $f(x)=p(1-p)^{x-1}, x=1,2,\cdots$

(2) 期望值為 $E(X)=\dfrac{1}{p}$

(3) 變異數為 $Var(X)=\dfrac{1-p}{p^2}$

(4) 幾何分佈有無記憶性質：$P(X>s+t\mid X>s)=P(X>t), s,t=1,2,\cdots$

例：投擲一銅板，直至出現正面才停止，若已投了 s 次皆未出現正面（$X>s$），則再投擲 t 次仍未出現正面的機率（$X>s+t$），與重新投擲 t 次未出現正面（$X>t$）之機率相同。

證明：

$$
\begin{aligned}
P(X>s+t\mid X>s) &= \frac{P(X>s+t, X>s)}{P(X>s)} \\
&= \frac{P(X>s+t)}{P(X>s)} \\
&= \frac{(1-p)^{s+t}}{(1-p)^{s}} \\
&= (1-p)^{t} = P(X>t)
\end{aligned}
$$

DOI:10.6814/NCCU202100266

2.1.4 負二項分布(negative binomial distribution o 或巴斯卡分布 Pascal distribution）

給定一正整數r，持續進行伯努利試驗，直至r次成功才停止，設總共之試驗次數X，則稱有參數r及p之負二項分布，以$NB(r,p)$表示。若$X=x$，最後一次為成功，第一次至第$x-1$次中，有$r-1$次成功，$x-r$次失敗，故

(1) 機率質量函數為$P(X=x)=C_{r-1}^{x-1}p^r(1-p)^{x-r}=C_{x-r}^{x-1}p^r(1-p)^{x-r}$ ， $x=r,r+1,\cdots$

(2) 期望值為$E(X)=\dfrac{r(1-p)}{p}$

(3) 變異數$Var(X)=\dfrac{r(1-p)}{p^2}$

比較：二項分布為投擲次數固定，看出現幾次正面

負二項分布為要求出現正面次數（r）固定，看要投擲幾次（X）

※在負二項分布中，令$r=1$便得幾何分布。

※設X有$NB(r,p)$分布，令$Y=X-r$，則Y表得到第r次成功停止時總共失敗之次數，且$P(Y=y)=P(X=r+y)=C_y^{r+y-1}p^r(1-p)^y$ ， $y=0,1,2,\cdots$

2.2 與時間有關的連續型分布

2.2.1 指數分佈(Exponential distribution)

若令W表示在 Poisson 過程中，由開始到第一次事件發生的時間為隨機變數，則

$$P(W>t)=P(\text{在}[0,t]\text{中無事件發生})$$

$$=e^{-\lambda t} \quad ,t>0$$

DOI:10.6814/NCCU202100266

又 $F_W(t)=P(W\le t)=1-P(W>t)=1-e^{-\lambda t}$

$$\Rightarrow f_W(t)=F_W'(t)=\lambda e^{-\lambda t}\ ,t>0$$

即 $f_W(t)=\lambda e^{-\lambda t}\ ,t>0$ 為指數分布的機率質量函數。指數分布為連續型的分布，其統計性質如下：

(1) 機率質量函數為 $f\left(x\right)=\lambda e^{-\lambda x}\ ,x>0$

(2) 期望值為 $E\left(X\right)=\dfrac{1}{\lambda}$

(3) 變異數為 $Var\left(X\right)=\dfrac{1}{\lambda^2}$

2.2.2 布朗運動與一維隨機行走(Brownian Motion)

布朗運動(Brownian motion)又稱 Wiener 過程，為最早被徹底研究的一個過程，與 Poisson 過程同為應用機率中最重要的兩個過程。其在 1827 年由英國科學家以顯微鏡觀察懸浮於水中的花粉粒子，發現這些花粉粒會做連續快速而不規則的隨機移動，且具有以下特性：

(1) 粒子的運動為平移與轉移，軌跡幾乎沒有切線。

(2) 粒子移動具有為獨立性。

(3) 溫度高低影響粒子的活潑性。

(4) 粒子的成分與密度與運動無相關性。

(5) 粒子運動永不停止。

假想一個粒子在水平直線上一步步的左右移動，且每步距離皆為一個單位長，粒子向右移動與向左移動的機率各為 p 與 $q=1-p\quad(0\le p,q\le 1)$，假設每一單位時間只移動一步(向右為+1，向左為-1)而且第 n 步在地 n 個瞬間獨立動

DOI:10.6814/NCCU202100266

作，令 X_n 為粒子第 n 步的位移，則獨立變數 X_n 為集合 $\{+1,-1\}$ 的取值，且滿足 $P(X_n=1)=p$ ，$P(X_n=-1)=q$ ，今以 W_0 表粒子的原始位置，則在時間 n 時粒子的位置表為 $W_n=W_0+X_1+X_2+\cdots+X_n$ ，此一序列的隨機變數 $W_n, n=1,2,3,\cdots$ 稱為隨機漫步。

2.2.3 常態分佈(高斯分佈)(Normal distribution)

常態分佈為連續型的分布，其圖形為鐘形曲線且具對稱性。中央極限定理使得常態分佈可當許多大樣本的近似分佈。常態分佈有兩個參數 μ 及 σ^2 ($\mu\in\Re, \sigma>0$)，以 $X\sim N\left(\mu,\sigma^2\right)$ 表示。μ 為位置參數，影響圖形位置，σ 為尺度參數。以下為常態分布的統計性質：

(1) 機率質量函數為 $f\left(x\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x\in\Re$

(2) 期望值為 $E(X)=\mu$

(3) 變異數為 $Var(X)=\sigma^2$

(4) 若 X 有 $N\left(\mu,\sigma^2\right)$ 分佈，則 $Z=\frac{X-\mu}{\sigma}$ 有 $N\left(0,1\right)$ 分佈，稱為標準常態分佈 (standard normal)，Z 的機率質量函數為 $f\left(x\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, x\in\Re$

常態分佈的機率值皆可藉由標準常態分佈之機率值而得，常令

$\Phi\left(x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du, x\in\Re$ 表標準常態分佈之分佈函數，一般常態分佈可先利用下式轉換成標準常態分佈

DOI:10.6814/NCCU202100266

$$
\begin{aligned}
P\left(a < X \le b\right) &= P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}) \\
&= P(Z \le \frac{b-\mu}{\sigma}) - P(Z \le \frac{a-\mu}{\sigma}) \\
&= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})
\end{aligned}
$$

而求得機率值 $\Rightarrow \Phi(z_p) = P(Z \le z_p) = p$

其中 Z 的期望值為 $E(Z)=0$，變異數為 $Var(Z)=1$

2.2.4 萊維穩定分布(Lévy Stable Distribution)

萊維穩定分布為連續型的機率分布，穩定分布由特徵函數 $\varphi(k)$ 及四個參數：特徵指數 α、偏態系數 β、尺度因子 c、平均值 μ 表達，其中 $0<\alpha\le 2$，$-1\le\beta\le 1$，$c>0$，$\mu>0$，萊維穩定分布有三種表示形式：

(1) $\alpha=\frac{1}{2}$，$\beta=1$

(2) $\alpha=1$，$\beta=0$

(3) $\alpha=2$

其中 $\alpha=2$ 時為常態分佈，而 α 從 2 降到 0 時，分布的厚尾會越來越明顯，

其一般函數形式 $F(x)=\frac{1}{2\pi}\int\varphi(k)e^{-ixk}dk$，$\varphi(k)=Ae^{-c|k|^{\alpha}}$

DOI:10.6814/NCCU202100266

第三章 二位元序列中的連續列探討

在施[13]與吳[14]的分析中，定義上漲(或下跌)事件為收益超過一個門檻值，其持續超過門檻的時間視為該事件存在的時間，而計算高頻的收益時，不同的時間點所取的計算收益時間片段、彼此間會有重疊；因此在 1 的連續列(見附錄)計算時，我們以重疊計數(over-lapping)的方式為考量，如圖 3.1。此計算方式在二位元序列組合分析中有關連續成功所需的等待時間(waiting time)的計數亦有考量，本章即針對固定長度的二位元字串做 1 的連續列片段的統計分析，整理出相關的計算，主要內容參考文獻[15]和[17]。

圖 3.1：在連續列片段中，若計算連續 3 個 1 的重疊計數(over-lapping)，以此圖的連續列為例，則共有 5+2=7 個。

3.1 出現給定長度連續 1 所需的等待時間

我們先分析連續列第一次出現的機率問題，此問題談的正是第二章中的幾何分布；然後再進一步分析連續列第二次以後的出現機率，由於連續列的出現涉及到取樣方式的不同，所以有多種情況，我們採用參考資料[15]的分類方式，考慮三種型態的等待時間(waiting time)。

DOI:10.6814/NCCU202100266

3.1.1 第一次出現給定長度連續 1 所需的等待時間

當 $X_1, X_2, \cdots$ 為 0 與 1 的序列，且出現 0 與出現 1 為相同分布的獨立事件 (iid)，其出現 1 機率為 $p = P(X_i = 1)$，出現 0 機率為 $q = 1 - p = P(X_i = 0)$，隨機變數 T_k 為第一次觀察到連續 k 個 1 所需經過的字串長度，若 $k = 1$，則隨機變數 T_1 的分布為幾何分布（$T_1 \sim Geo(p)$），若 $k > 1$，則隨機變數 T_k 的分布稱為 k 階的幾何分布，因此令 $f(x) = P(T_k = x)$ 為 k 階幾何分布的機率質量函數，則在 $0 \le x \le 2k$ 時可得到

$$f(x) = \begin{cases} 0 & if\ 0 \le x < k \\ p^k & if\ x = k \\ qp^k & if\ k < x \le 2k \end{cases}$$

若 $x > 2k$ 時，則有 $f(x) = qp^k \left(1 - \sum_{i=0}^{x-k-1} f(i)\right)$ 的遞迴關係。此外，在 x 個等待的時間中，最後一片段為第一次連續 k 個 1 的 $\underbrace{11\cdots1}_{k}$，則其餘的 $x-k$ 的等待時間假設 0 有 x_1 個，10 有 x_2 個，110 有 x_3 個，⋯，$\underbrace{11\cdots1}_{k-1}0$ 有 x_k 個，則此 $x_1 + x_2 + \cdots + x_k$ 個片段的排列數為 $\frac{(x_1 + x_2 + \cdots + x_k)!}{x_1! x_2! \cdots x_k!}$，且 $\sum_{i=}^{k} i x_i + k = x$，其機率質量函數亦可表為

$$\begin{aligned} f(x) &= \sum \frac{(x_1 + x_2 + \cdots + x_k)!}{x_1! x_2! \cdots x_k!} q^{x_1} (pq)^{x_2} \cdots (p^{k-1} q)^{x_k} p^k \\ &= \sum \frac{(x_1 + x_2 + \cdots + x_k)!}{x_1! x_2! \cdots x_k!} q^{x_1 + x_2 + \cdots + x_k} p^{k + \sum_{i=1}^{k} (i-1) x_i} \\ &= \sum \frac{(x_1 + x_2 + \cdots + x_k)!}{x_1! x_2! \cdots x_k!} q^{\sum_{i=1}^{k} x_i} p^{k + \sum_{i=1}^{k} i x_i - \sum_{i=1}^{k} x_i} \\ &= \sum \frac{(x_1 + x_2 + \cdots + x_k)!}{x_1! x_2! \cdots x_k!} q^{\sum_{i=1}^{k} x_i} p^{x - \sum_{i=1}^{k} x_i} \\ &= \sum \frac{(x_1 + x_2 + \cdots + x_k)!}{x_1! x_2! \cdots x_k!} p^x \left(\frac{q}{p}\right)^{x_1 + x_2 + \cdots + x_k} \end{aligned}$$

DOI:10.6814/NCCU202100266

3.1.2 第 r 次出現給定長度連續 1 所需的等待時間

接下來我們考慮連續列第 r 次（$r>1$）出現的情況，若 $T_k^{(1)},T_k^{(2)},\cdots,T_k^{(r)}$ 為獨立的隨機變數且其分布皆為 k 階的幾何分布，第二章中所談的負二項分布為幾何分布的衍伸，在此我們考慮隨機變數為 $T_k^{(1)}+T_k^{(2)}+\cdots+T_k^{(r)}$，由於取樣方式的不同，其分布需要做進一步的細分，我們採用三種不同的取樣方式[15]：

(1) 第一型等待時間分布：

為非重疊計數(non-overlapping counting)，一旦出現連續 k 個 1，計算出連續 k 個 1 的等待時間，並且接著重新開始計數，此種有關第 r 個片段為連續 k 個 1 的字串長度記為 $T_{r,k}^{(I)}$。設前 15 次結果的序列為 110111001111101，則 $T_{1,2}^{(I)}=2$，$T_{2,2}^{(I)}=5$，$T_{3,2}^{(I)}=10$，$T_{4,2}^{(I)}=12$，$T_{5,2}^{(I)}>15$，同樣的，也具有 $T_{1,3}^{(I)}=6$，$T_{2,3}^{(I)}=11$，$T_{3,3}^{(I)}>15$。

隨機變數 $T_{r,k}^{(I)}$ 的機率質量函數為

$$
\begin{aligned}
f_{r,k}^{(I)}(x) &= P(T_{r,k}^{(I)}=x) \\
&= \sum \frac{(x_1+x_2+\cdots+x_k+r-1)!}{x_1!x_2!\cdots x_k!(r-1)!} p^x \left(\frac{q}{p}\right)^{x_1+x_2+\cdots+x_k}
\end{aligned}
$$

(2) 第二型等待時間分布：

連續 k 個 1 或連續更多 1 視為一次的等待時間，不管此字串多長。則此種第 r 個連續 1 的等待時間方案被記為 $T_{r,k}^{(II)}$。設前 15 次結果的序列為

DOI:10.6814/NCCU202100266

110111001111101，則 $T_{1,2}^{(II)}=2$ ，$T_{2,2}^{(II)}=5$ ，$T_{3,2}^{(II)}=10$ ，$T_{4,2}^{(II)}>15$ ，，同樣的，也具有 $T_{1,3}^{(II)}=6$ ， $T_{2,3}^{(II)}=11$ ，$T_{3,3}^{(II)}>15$ 。

(3) 第三型等待時間分布：

為重疊計數(overlapping counting)一個 $l\geq k$ 個 1 且不間斷的連續列，計算次數為 $l-k+1$ 個連續 k 個 1 的運行，則此種第 r 個連續 1 重疊運行的等待時間被記為 $T_{r,k}^{(III)}$ 。設前 15 次結果的序列為 110111001111101，則 $T_{1,2}^{(III)}=2$ ，$T_{2,2}^{(III)}=5$ ，$T_{3,2}^{(III)}=6$ ，$T_{4,2}^{(III)}=10$ ，$T_{5,2}^{(III)}=11$ ，$T_{6,2}^{(III)}=12$ ，$T_{7,2}^{(III)}=13$ ，$T_{8,2}^{(III)}>15$ ，同樣的，也具有 $T_{1,3}^{(III)}=6$ ，$T_{2,3}^{(III)}=11$ ，$T_{3,3}^{(III)}=12$ ，$T_{4,3}^{(III)}=13$ ，$T_{5,3}^{(III)}>15$ 。

3.2 樣本空間的分析與模擬

在以隨機產生二位元序列做為模擬前，我們考慮一個集合，包含總長度為 n 的所有二位元字串，則元素總數為 2^n 個。在施[13]與吳[14]的研究工作中，計算上漲事件中的持續時間分布，在此我們考慮以事件為基礎的樣本空間，分別以完整的連續列的樣本空間 U ，與次連續列為單位的樣本空間 U_s ([13]與[14]的研究中的計算方式)。

DOI:10.6814/NCCU202100266

3.2.1 完整連續列的樣本空間

首先我們考慮完整連續列的樣本空間U，不論連續列的長短，其出現的機會都一樣，其機率由2^n個序列中出現的數目來決定。因此我們計算長度為L的連續列數目(以$N_{L=n-k}$表示[17])，樣本空間U(包含所有L，$1 \le L \le n-1$)的元素總數為(詳細計算參考附錄A)$n(U)=\sum_{L=1}^{n-1} N_{L=n-k} = \sum_{k=1}^{n-1}(k+3)\cdot 2^{k-2}$ 。

在樣本空間中U，長度為L的事件具下列性質(詳細推導見附錄B):

(1) 機率質量函數為 $P(L)=\dfrac{(n-L+3)\cdot 2^{n-L-2}}{(n+1)\cdot 2^{n-2}-1}$ ， $1 \le L \le n-1$ (3.1)

(2) 期望值 $E(L)=\dfrac{n\cdot(2^{n-1}-1)}{(n+1)\cdot 2^{n-2}-1}$ ， $1 \le L \le n-1$ (3.2)

(3) 變異數

$$Var(L)=\frac{(n^2-n-4)\cdot 2^{2n-3}-(n^3-3n^2+4n-10)\cdot 2^{n-2}-2}{[(n+1)\cdot 2^{n-2}-1]^2} \quad , \quad 1 \le L \le n-1 \qquad (3.3)$$

值得注意的是，當序列總長度$n \to \infty$時，機率質量函數$P(L) \approx 2^{-L} = e^{-(\log 2)L}$，為一指數函數。

3.2.2 次連續列(subsequence)的樣本空間

其次我們考慮次連續列(subsequence)的樣本空間U_s，在長度為$L \ge s$的連續列中，次連續列長度為$s\,(1 \le s \le L)$的出現次數為重疊計數(如圖 3.1)，且不論L和s為何，次連續列皆視為一個事件，各出現的機會都一樣。例如$n=8$時，所有的次連續列(即所有事件)其出現的機率列於表 3.1。

DOI:10.6814/NCCU202100266

表 3.1

在 2^n 個可能的字串(string)逐一檢查：$n=8$					
k	$n-k$（L）	長度為 L 的連續列 (run of 1's) 個數	隨機選取連續列, 其長度為 L 的機率	長度為 s 的次連續列 (sub run of 1's) 個數	隨機選取長度為 s 的次連續列, 持續時間 $T=L-s+1$ 其出現的機率
1	7	2	0.003478261	10=2*(7-3+1)	0.053763441($T=5$)
2	6	5	0.008695652	20	0.107526882($T=4$)
3	5	12	0.020869565	36	0.193548387($T=3$)
4	4	28	0.048695652	56	0.301075269($T=2$)
5	3	64	0.111304348	64	0.344086022($T=1$)
6	2	144	0.250434783		
7	1	320	0.556521739		
註		和=575	和=1	和=186	和=1

其中表 3.1 中最後一欄中所列的 τ 為給定長度 L 的連續片段，其中長度為 s 的次連續列連續存在的次數 $T=L-s+1$ 相當於施[13]與吳[14]討論中的持續時間。

在一般情況下，持續次數 $T(L,s)$ 的機率質量函數(推導見附錄 C)為

$$P(T,s)=\frac{(n-T-s+4)\cdot 2^{n-T-s-1}\cdot T}{(n-s+1)[2^{n-s}-1]} \tag{3.4}$$

在此公式中，我們將 L 視為 T 和 s 的函數。不管 L 為何，長度為 s 的次連續列，持續存在 T 步的機率藉由(3.4)式來描述。

DOI:10.6814/NCCU202100266

3.2.3 電腦模擬:以單一的長序列產生樣本空間中短序列連續 1(run of ones)的分布

我們將二位元字串的總長度擴增到 $n=1000$，並以電腦樣本在 0 與 1 出現機會均等下產生 $n=1000$ 的二位元字串，在所有連續列中，以重疊計數(over-lapping)的計數方式統計次連續列長度 $s=3$ 所有的連續列中的出現次數，橫軸為持續時間 T，縱軸為 T 出現的機率。與(3.4)式比較下，是吻合的。

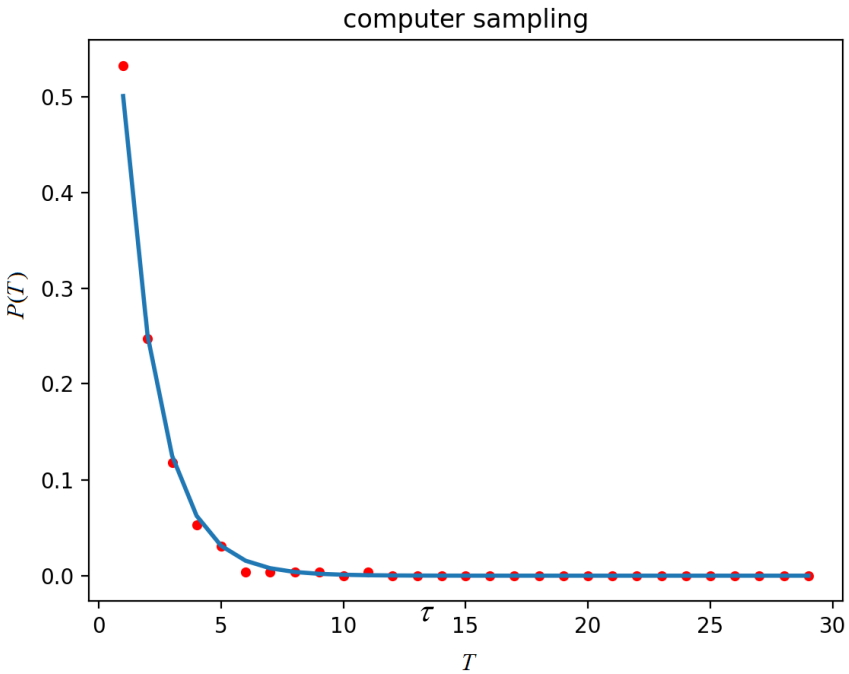


圖 3.2：此為固定長度為 $s=3$ 的次連續列其持續時間 T 的機率密度函數。紅色點為電腦隨機產生 $n=1000$ 的二位元樣本計算結果，藍色曲線為(3.4)式的公式。

(3.4)式背後假設為 0 與 1 出現的機會一樣的 2^n 個長度為 n 的序列構成的樣本空間，而此處藉由為所檢驗的次連續列長度($s=3$)將近三個數量級長度($n=1000$)的單一序列的模擬計算，就足以產生和公式一致的結果。若所取 s 增

DOI:10.6814/NCCU202100266

大的話，所需選取的單一序列長度便需要加長。

此處分析(3.4)式當序列總長度$n \to \infty$時，$P(T,s) \approx T \cdot 2^{-T-1}$，隨$T$變大，主要為指數函數遞減的形式。隨機產生的序列其持續時間的機率分布基本上為指數函數，則冪次律的產生需要由非隨機產生的序列做為探討。

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100266

第四章 二位元序列的隨機過程與記憶效應

在第三章我們觀察到隨機產生的序列其持續時間的機率分布基本上為指數函數。在本章中我們將電腦產生的序列視為一個時間的序列，在某一時間點的下一個位元(股價變動)，可以是任意產生(交易不依靠記憶)或是根據股價變動的歷史來決定(根據記憶進行股價變動)，因此在隨機產生二位元的序列中，每一步新位元的產生的方式如下:統計前 m 步中 1 的個數，以作為新產生 1 的機率。在無記憶($m=0$)時，隨機產生二位元的序列在樣本空間中為浮動的狀態；在有記憶($m>0$)時，系統會收斂到在樣本空間中的固定點。以下我們就記憶長度對所產生時間序列的影響的做一番探討。

我們先探討記憶的效應是否會改變連續列長度分布的函數形式(由指數函數到冪次律)，我們考慮記憶的作用是一個簡單的形式，在此情況下答案是否定的(第 4.1 節)。而我們觀察到，記憶的範圍是否涵蓋所有的歷史，會造成系統具有非常不一樣的收斂結果，其收斂形式為本章計算模擬探討的主要議題。在 4.2 節中我們探討完全記憶的情況(所有的過去歷史皆納入決定下一步的考量)；4.3 節我們追蹤有限記憶的情況下，系統收斂的情形。

DOI:10.6814/NCCU202100266

4.1 固定記憶長度 m 的隨機過程，連續列的分布過程為指數函數

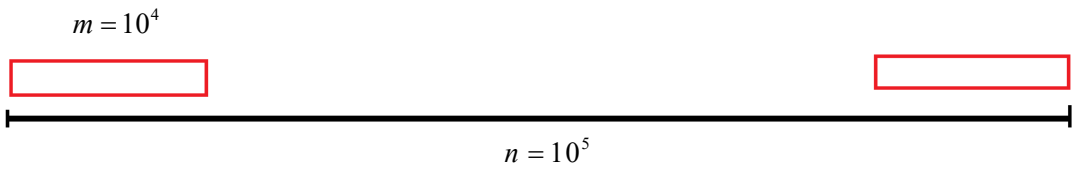


圖 4.1：在固定記憶長度 m 下，隨機產生位元序列的步驟示意圖。統計前 m 步中 1 的比例，以作為新產生 1 的機率，視窗的移動(scan)方法為每產生一個新的位元，便捨去記憶視窗中最前面的位元，使其記憶長度固定為 m 。

在股市中，股價的漲跌會受交易歷史的影響。因此我們先討論依記憶視窗中 0 與 1 的比例決定下一位元的機制。我們考慮記憶長度為 $m=10^4$，即為先以 $\frac{1}{2}$ 比例隨機產生 0(或 1)得出第一個起始視窗的字串，再統計起始視窗中的 0 與 1 的個數，作為下一筆產生的機率依據，例如:若起始視窗以電腦隨機模擬中有 4996 個 1，5004 個 0，則以 0.4996 和 0.5004 為產出下一筆 1 和 0 的機率，得出第 10001 位元資料後，若為 1，統計當前第 2 筆至第 10001 筆(共 10^4 筆)的 0 與 1 個數，若有 4997 個 1，5003 個 0，則以 0.4997 和 0.5003 分別為產出下一筆 1 和 0 的機率…，假設全部的模擬共有 10^5 步，如圖 4.1 所示，系統演進的過程如同一個長度為 10^4 視窗的移動。圖 4.2 為最後一個視窗中的統計結果，圖 4.3 則為對其縱軸取對數，發現此連續列分布為指數遞減型式。

DOI:10.6814/NCCU202100266

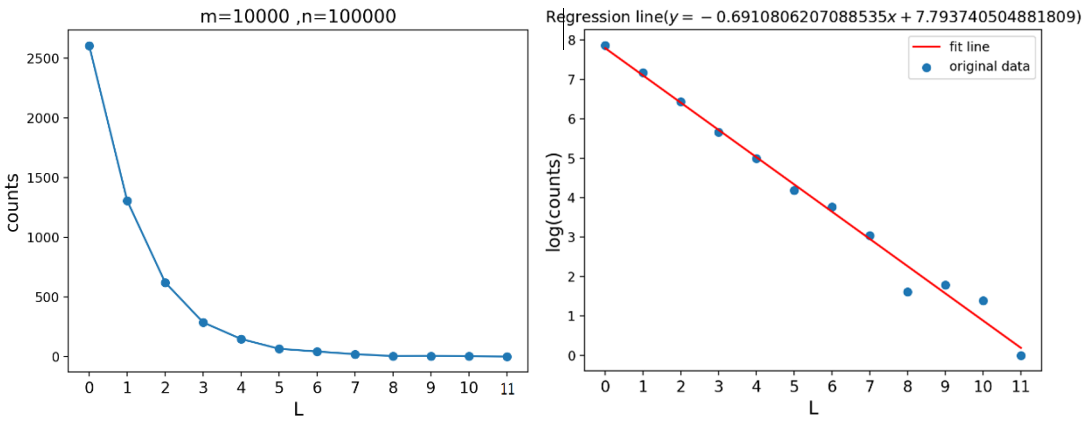


圖 4.2：左圖為依圖 4.1 的程序產生 10^5 二位元序列，其連續列的個數分布圖。右圖為縱軸為對數軸。可以看出曲線為指數函數。

4.2 全記憶的模擬:立即更新與變動更新

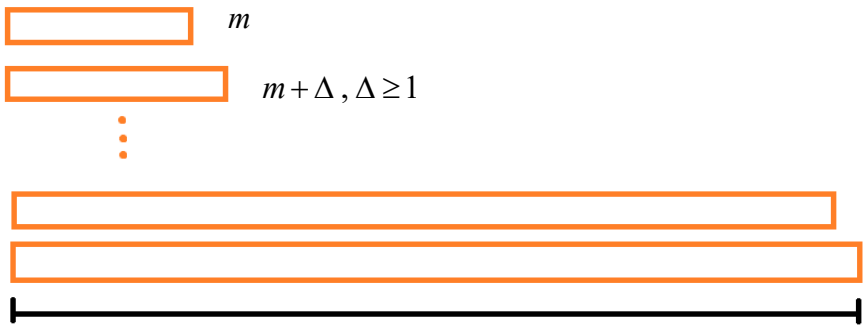


圖 4.3：在全記憶條件下，隨機產生位元序列的步驟示意圖。相較於圖 4.1 的固定記憶長度的移動方式，此處的記憶長度持續擴增，擴增時機詳見本文說明。

DOI:10.6814/NCCU202100266

本節探討記憶範圍涵蓋所有的歷史，系統演進的情況(圖 4.3)。我們把記憶更新方式分為以下兩種情況：

(1) 立即更新:隨時間演進隨時更新記憶，因此視窗每次擴張 1 個位元。

(2) 變動更新:進行中連續 1 的序列終止時才進行記憶更新。

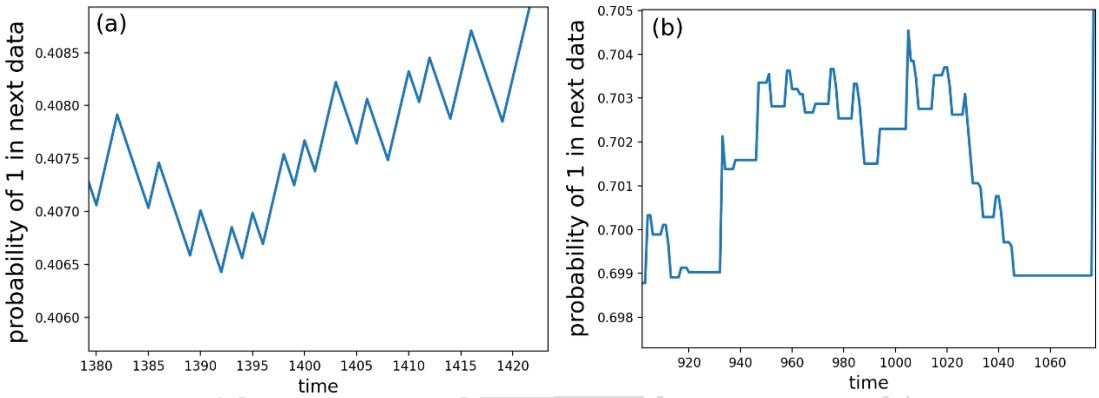


圖 4.4：(a)立即更新(b)變動更新，記憶視窗中的時間演進圖。詳細的模擬細節分別見圖 4.5 和圖 4.6。

以下我們就此兩種不同狀況的記憶更新做為討論：

(1) 立即更新(圖 4.3 中 $\Delta=1$):

先以 $\frac{1}{2}$ 比例隨機產生 0(或 1)得出第一個起始視窗的字串，再統計長度為 8 的起始視窗中的 0 與 1 的個數，作為下一筆產生的機率依據，例如:若起始視窗以電腦隨機模擬中有 3 個 1，5 個 0，則以 $\frac{3}{8}$ 和 $\frac{5}{8}$ 為產出下一筆 1 或 0 的機率，得出第 9 筆資料後，若為 1，統計當前共 9 筆的 0 與 1 個數，共有 4 個 1，5 個 0，則以 $\frac{4}{9}$ 和 $\frac{5}{9}$ 為產出下一筆 1 或 0 的機率，依此程序得到 $n=10000$ 的字串，並觀察每一步記憶視窗中 1 所佔的比率，結果見圖 4.5。由圖可見系統一開始 1 出現的機率有比較大的震盪，接著隨記憶視窗的擴張呈現穩定，直到模擬終結。

DOI:10.6814/NCCU202100266

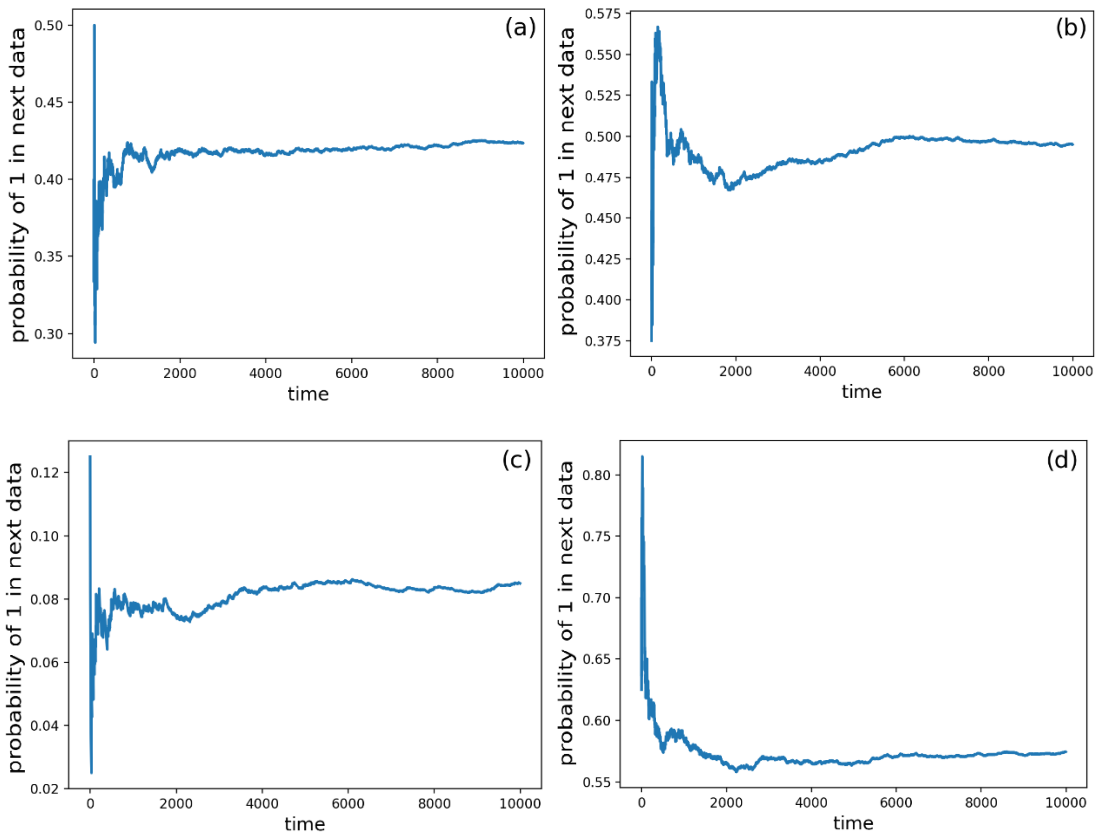


圖 4.5：記憶長度隨時間立即更新所得到 1 的機率時間演進圖。模擬的方式為以 $\frac{1}{2}$ 為產生 0(或 1)的機率隨機產生長度為 8 的字串，統計當下觀察到的歷史紀錄，其中 0 與 1 的個數產生更新下一筆的機率，橫軸為步數(時間)，縱軸為決定下一步位元為 1 的機率。其中(a)的放大圖見圖 4.4(a)。

(2) 變動更新(圖 4.4 中 $\Delta \geq 1$):

有連續 1 片段終止時才做視窗更新擴張，同立即更新的模擬，仍然以 $\frac{1}{2}$ 比例隨機產生 0(或 1)得出第一個起始視窗的字串，再統計長度為 8 的起始視窗中 0 與 1 的個數，作為下一筆產生的機率依據，例如:若起始視窗以電腦隨機模擬中有 3 個 1，5 個 0，則以 $\frac{3}{8}$ 和 $\frac{5}{8}$ 為產出下一筆 1 或 0 的機率，得出第 9 筆資料

DOI:10.6814/NCCU202100266

後，分為以下兩種狀況：

①若為 1 則繼續以$\frac{3}{8}$和$\frac{5}{8}$為產出下一筆 1 或 0 的機率，直到得出下一筆為 0 再重新統計當前視窗的 1 與 0 個數更新作為下一筆產出 1 或 0 的機率。

②若為 0 則繼續以$\frac{3}{8}$和$\frac{5}{8}$為產出下一筆 1 或 0 的機率，直到得出下一筆為①的狀況。

經過$n=10000$步，我們觀察到無論視窗擴增大小都會得出最後記憶視窗中，1 的個數在視窗中所占的比率趨於穩定，見圖 4.6。

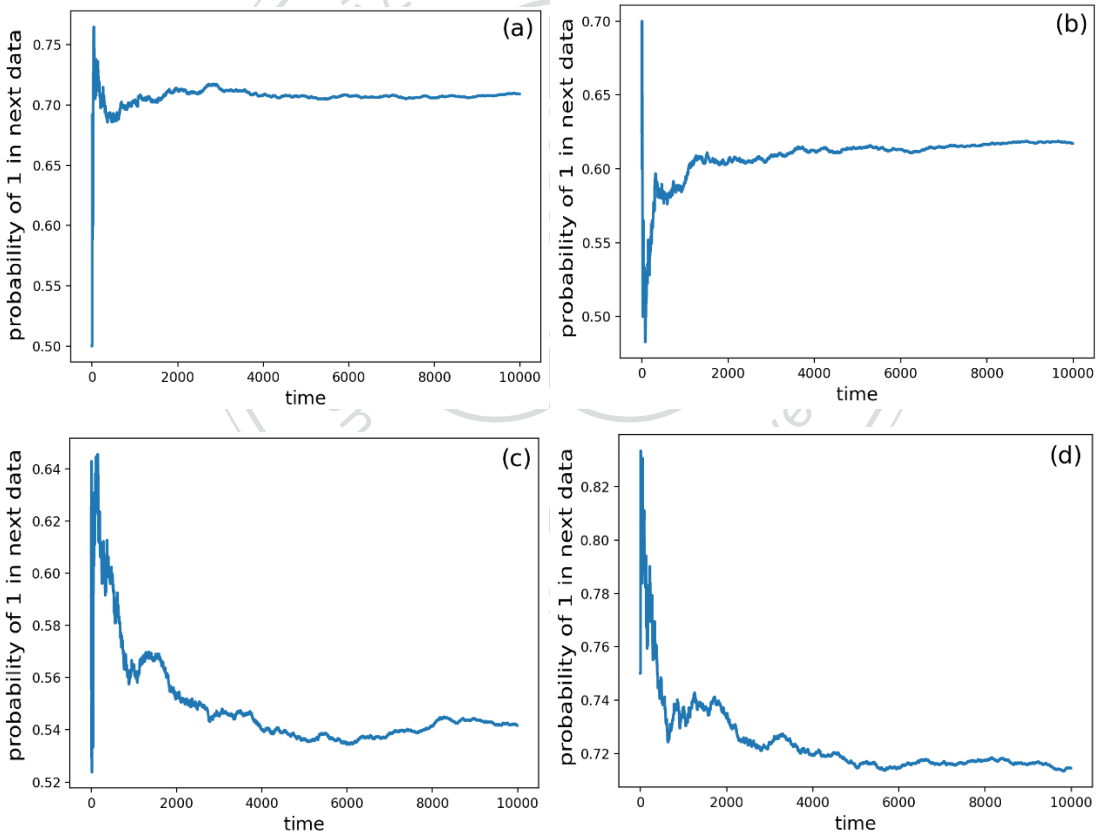


圖 4.6：同圖 4.5，為變動更新的情況。此處記憶視窗更新為進行中連續 1 的序列終止時才進行記憶更新。其中(a)的放大圖見圖 4.4(b)。

DOI:10.6814/NCCU202100266

4.3 有限記憶($m>1$ 且固定)二位元序列的長時間收斂分析

在 4.1 節的模擬當中，固定記憶長度下的連續列分布為指數形式，而在 4.2 節的全記憶模擬中系統狀態會趨於穩定。而對於固定記憶長度的系統中，我們可以預期它會收斂到全為 0 或全為 1 的固定狀態(記憶視窗中全為 0 或全為 1)。本節我們對記憶長度固定為 $m=2^i(i=3,4,5,6,7)$ 的系統分別做 1000 次的實驗，進行統計分析最後收斂到皆為 0 或皆為 1 的時間。

4.3.1 有限記憶二位元序列的時間演進

我們將模擬記憶長度固定為 $m=2^i(i=3,4,5,6,7)$，分別得到其在移動過程中，最後會收斂到視窗內皆為 0 或皆為 1 的時間，典型的實驗如圖 4.7 ($m=8$)、圖 4.8($m=16$)、圖 4.9($m=32$)、圖 4.10($m=64$)及圖 4.11 ($m=128$)。模擬的方式為以 $\frac{1}{2}$ 為產生 0(或 1)的機率隨機產生 m 個位元的字串為起始，統計當下的歷史記憶視窗，其中 0 與 1 的個數為更新下一筆的機率。模擬實驗相當於一個視窗的移動過程(如圖 4.1)最後抵達皆為 0 或皆為 1 的狀態，所需時間為 $T_{arrival}$。將各別記憶長度做 1000 次的模擬結果，統計其收斂的時間分布，列於圖 4.12。如圖所示，記憶視窗越寬，所需的收斂時間越長。

DOI:10.6814/NCCU202100266

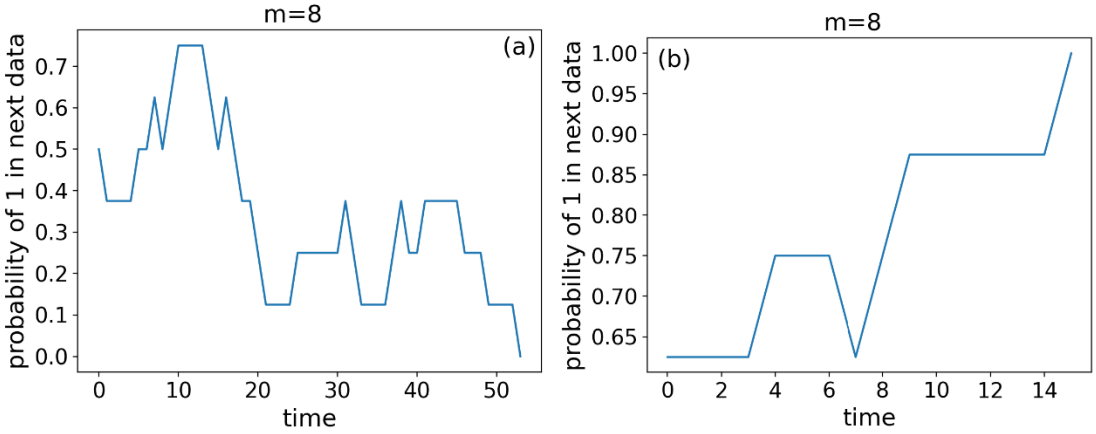


圖 4.7：單一模擬中(隨意選取的兩個例子)，記憶長度固定為 $m=8$ 時出現 1 的機率之時間演進圖。模擬的方式為以 $\frac{1}{2}$ 為產生 0(或 1)的機率隨機產生 8 個位元的字串為起始，統計當下 $m=8$ 的歷史記憶視窗，其中 0 與 1 的個數為更新下一筆的機率，橫軸為步數(時間)，縱軸為決定下一步位元為 1 的機率。由此圖可見在此一模擬中，(a)和(b)分別為系統在第 53 步和第 15 步時到達收斂點，亦即記憶視窗全為 0((a)的情形)或全為 1((b)的情形)。

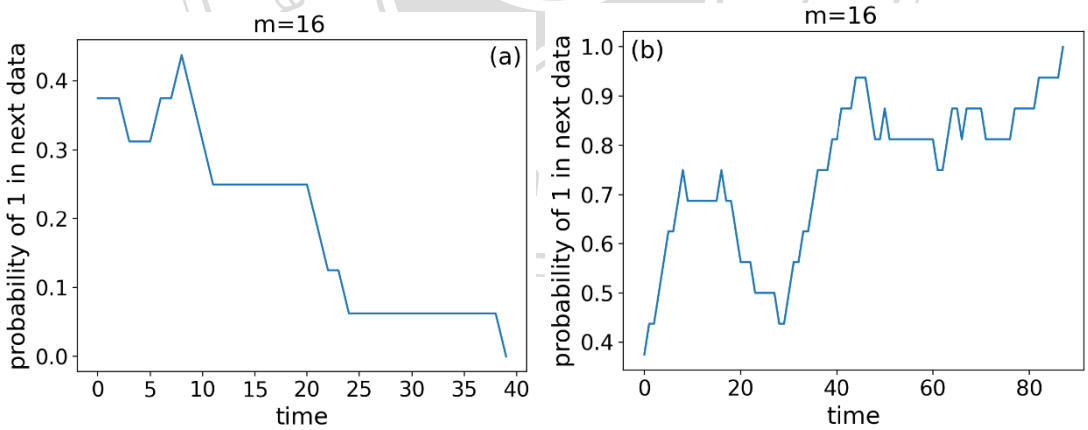


圖 4.8：同圖 4.7，記憶長度固定為 $m=16$ 時出現 1 的機率之時間演進圖。由此圖可見在此一模擬中，(a)和(b)分別為系統在第 39 步和第 87 步時到達收斂點，亦即記憶視窗全為 0((a)的情形)或全為 1((b)的情形)。

DOI:10.6814/NCCU202100266

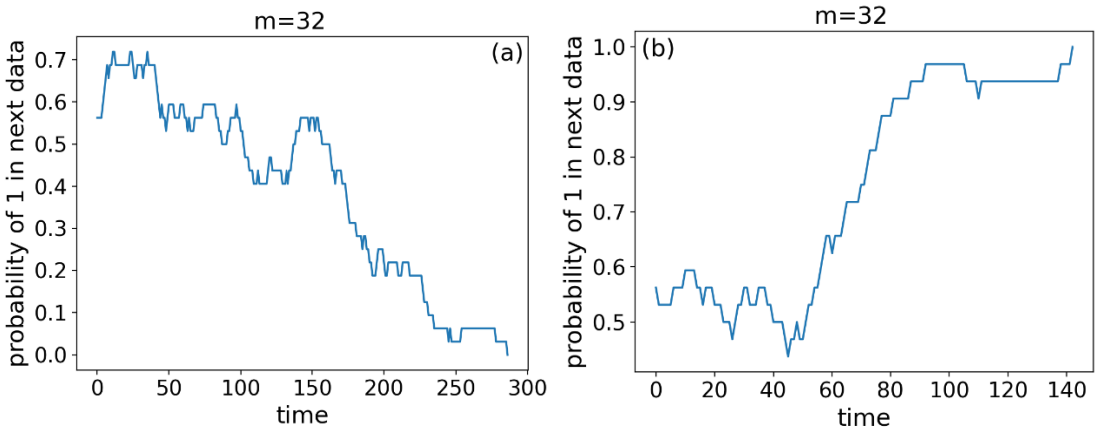


圖 4.9：同圖 4.7，記憶長度固定為 $m = 32$ 時出現 1 的機率之時間演進圖。由此圖可見在此一模擬中，(a)和(b)分別為系統在第 286 步和第 142 步時到達收斂點，亦即記憶視窗全為 0((a)的情形)或全為 1((b)的情形)。

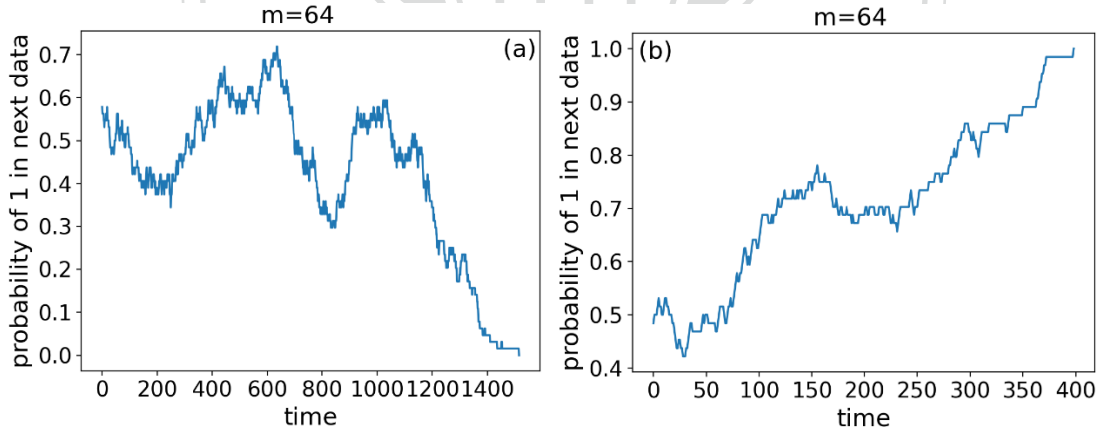


圖 4.10：同圖 4.7，記憶長度固定為 $m = 64$ 時出現 1 的機率之時間演進圖。由此圖可見在此一模擬中，(a)和(b)分別為系統在第 1515 步和第 398 步時到達收斂點，亦即記憶視窗全為 0((a)的情形)或全為 1((b)的情形)。

DOI:10.6814/NCCU202100266

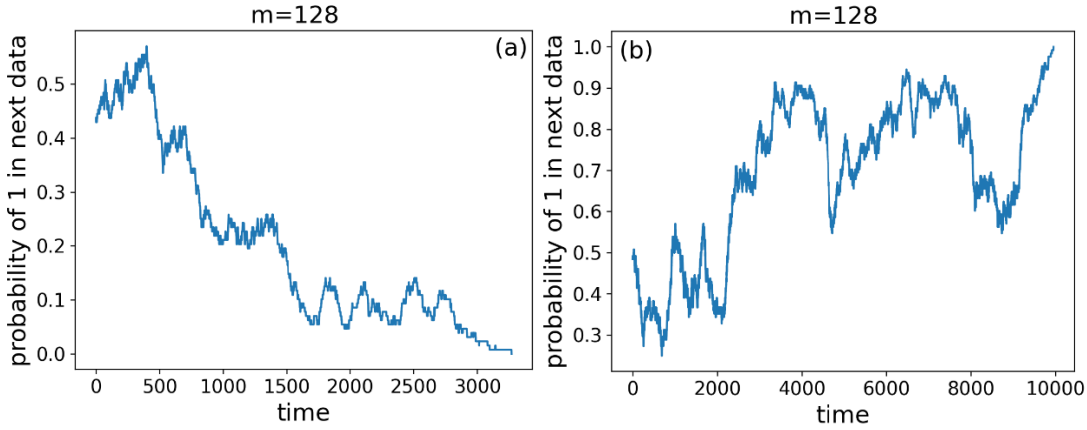


圖 4.11：同圖 4.7，記憶長度固定為 $m=128$ 時出現 1 的機率之時間演進圖。由此圖可見在此一模擬中，(a)和(b)分別為系統在第 3268 步和第 9946 步時到達收斂點，亦即記憶視窗全為 0((a)的情形)或全為 1((b)的情形)。

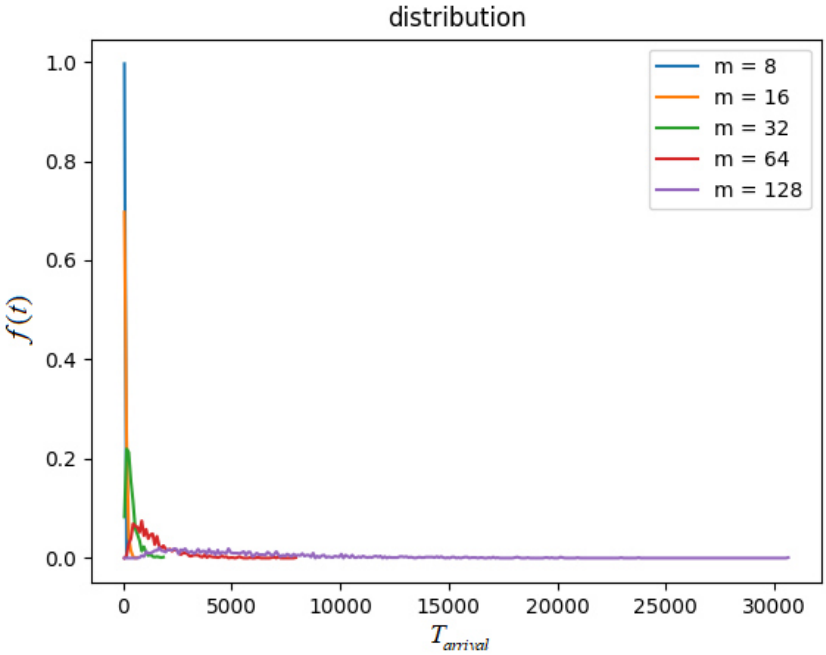


圖 4.12：固定記憶長度的收斂時間分布圖。我們將模擬記憶長度固定為 $m=2^i(i=3,4,5,6,7)$ 分別進行 1000 次的實驗，模擬實驗相當於一個視窗的移動過程(如圖 4.1)最後收斂抵達皆為 0 或皆為 1 的狀態，我們統計所需的時間。

DOI:10.6814/NCCU202100266

4.3.2 固定記憶二位元序列的隨機過程分析：轉移矩陣(transition matrix)

在此模擬中，我們分析固定記憶長度下的狀態，記憶視窗共可分為$m+1$種狀態：0 個 1(m 個 0)，1 個 1($m-1$個 0)，2 個 1($m-2$個 0)，3 個 1($m-3$個 0)，…，m 個 1(0 個 0)共$m+1$種狀態，其中若有i個 1，則出現 1 的機率為$\frac{i}{m}$，出現 0 的機率為$\frac{m-i}{m}$，而狀態間的轉移機率分為底下三種:

(1) 由i個 1 變為$i+1$個 1，為新增加一個 1，同時隨著視窗的移動，最前面丟掉一個 0，即 $\leftarrow \boxed{0\cdots\cdots} \leftarrow 1$ ，其機率為$\frac{m-i}{m}\cdot\frac{i}{m}$。

(2) 由i個 1 變為$i-1$個 1，為新增加一個 0，同時隨著視窗的移動，最前面丟掉一個 1，即 $\leftarrow \boxed{1\cdots\cdots} \leftarrow 0$ ，其機率為$\frac{i}{m}\cdot\frac{m-i}{m}$。

(3) 由i個 1 變為i個 1，為新增加一個 1，同時隨著視窗的移動，最前面丟掉一個 1，或為新增加一個 0，由於視窗的移動，最前面丟掉一個 0，即 $\leftarrow \boxed{1\cdots\cdots} \leftarrow 1$ 或 $\leftarrow \boxed{0\cdots\cdots} \leftarrow 0$ ，其機率為

$$\frac{i}{m}\cdot\frac{i}{m}+\frac{m-i}{m}\cdot\frac{m-i}{m}=\frac{(m-i)^2+i^2}{m^2}。$$

此過程的轉移狀態，我們可以建立$(m+1)$階的轉移方陣(transition matrix)

DOI:10.6814/NCCU202100266

$$
\boldsymbol{A}=\begin{pmatrix}
1 & \frac{m-1}{m^2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
0 & \frac{(m-1)^2+1}{m^2} & \frac{2(m-2)}{m^2} & & & & & & 0 \\
0 & \frac{m-1}{m^2} & \frac{(m-2)^2+4}{m^2} & \ddots & & & & & \\
\vdots & 0 & \frac{2(m-2)}{m^2} & \ddots & \ddots & & & & \vdots \\
 & & 0 & \ddots & 0 & & & & \\
\vdots & & & \ddots & \ddots & \frac{1}{2} & \ddots & & \vdots \\
 & & & & \ddots & \frac{1}{4} & \ddots & \ddots & \\
\vdots & & & & \ddots & \frac{1}{2} & \ddots & & \vdots \\
 & & & & & 0 & \ddots & 0 & \\
\vdots & & & & & \ddots & \frac{2(m-2)}{m^2} & 0 & \vdots \\
 & & & & & \ddots & \frac{(m-2)^2+4}{m^2} & \frac{m-1}{m^2} & 0 \\
0 & & & & & \ddots & \frac{2(m-2)}{m^2} & \frac{(m-1)^2+1}{m^2} & 0 \\
0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{m-1}{m^2} & 1
\end{pmatrix} \tag{4.1}
$$

此矩陣的兩個固定點為[1,0,0,…,0,0]和[0,0,0,…,0,1]，分別對應到全為 0 和全為 1 的記憶視窗序列，分析詳見附錄 D。

其中我們可以由數值計算 $\boldsymbol{A}^M$ 得到

$$
\lim_{M\to\infty}\boldsymbol{A}^M=\begin{pmatrix}
1 & \frac{m-1}{m} & \frac{m-2}{m} & \cdots & \frac{1}{2} & \cdots & \frac{2}{m} & \frac{1}{m} & 0 \\
0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 \\
\vdots & & & & & & & & \vdots \\
\vdots & & & & & & & & \vdots \\
\vdots & & & & & & & & \vdots \\
0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 \\
0 & \frac{1}{m} & \frac{2}{m} & \cdots & \frac{1}{2} & \cdots & \frac{m-2}{m} & \frac{m-1}{m} & 1
\end{pmatrix} \tag{4.2}
$$

DOI:10.6814/NCCU202100266

此結果顯示，由狀態 $\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix}$ 出發(此處 $a_0 + a_1 + \cdots + a_m = 1$，所描述的狀態為視窗中有 $m-j$ 個 1，j 個 0 的情況有 a_j 的機率)，若取 $a_j = 1$，其餘 $a_k = 0\,(k \neq j)$，此狀態為 m 位元視窗中，有 m - j 個 1，其餘為 0，則系統最後收斂結果為：有 $\frac{m-j}{m}$ 的機會是視窗全為 1，另外 $\frac{j}{m}$ 的機會是視窗全為 0。取 $m = 8$，矩陣(4.2)隨 M 增大收斂的數值計算及分析見附錄 E。

4.3.3 固定記憶二位元序列的隨機過程分析：$T_{arrival}$

我們進一步分析圖 4.12 的結果，我們發現收斂時間的平均值與記憶長度呈現平方的冪次關係，圖 4.13 為收斂平均時間對記憶長度的關係圖，圖 4.14 為擬合的結果，擬合得到的函數為

$$T_{arrival} = 2^{-1.1} \cdot m^2 \tag{4.3}$$

我們根據公式(4.3)的結果，對圖 4.12 的各條分布曲線做尺度變換

$$f^*(T_{arrival}^*) = f(T_{arrival}) \cdot m^2,\ T_{arrival}^* = \frac{T_{arrival}}{m^2} \tag{4.4}$$

得到的結果列於圖 4.15，可看出變換後的曲線基本上重合在一起。

DOI:10.6814/NCCU202100266

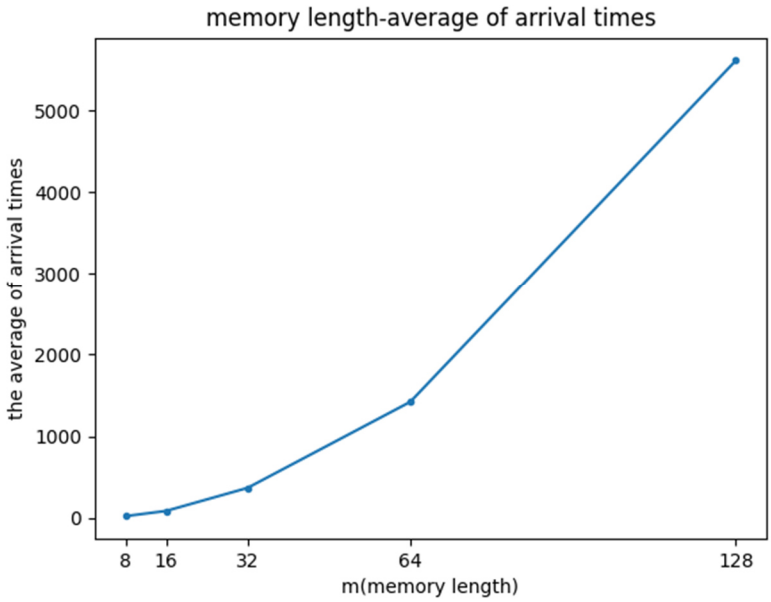


圖 4.13：收斂平均時間對記憶長度的關係圖。將圖 4.12 的分布圖各別計算平均值得到的結果。

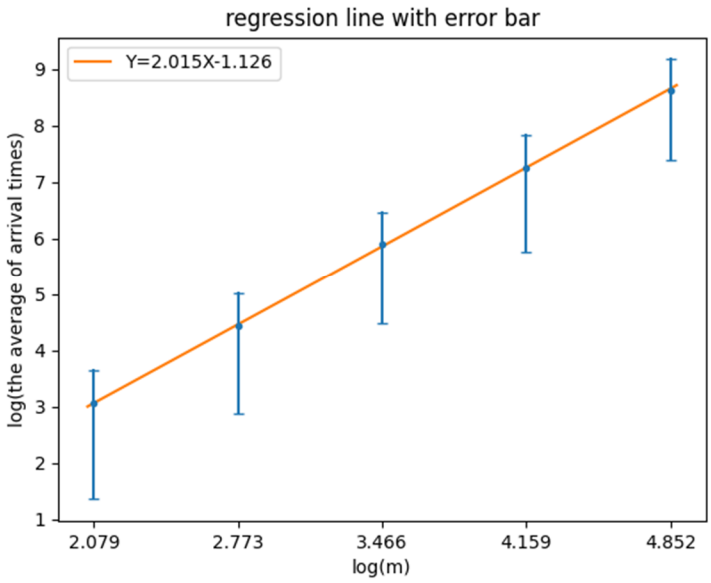


圖 4.14：收斂平均時間對記憶長度的數據做擬合的對數圖。數據來自圖 4.12 與圖 4.13。圖中所標示各數據點的有效範圍，是由 1000 個樣本的標準差所計算出。

DOI:10.6814/NCCU202100266

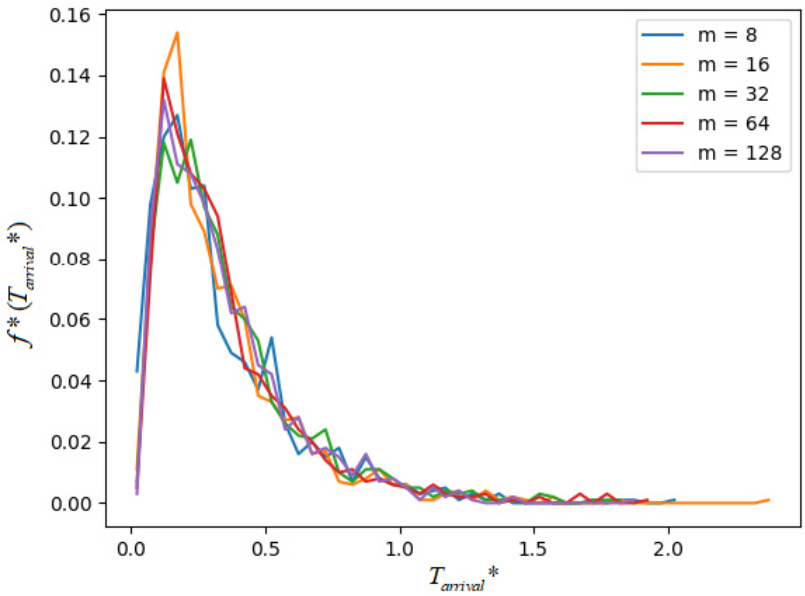


圖 4.15：尺度變換後，收斂時間對固定記憶長度的分布圖。我們對圖 4.12 的各條分布曲線做尺度變換(式(4.4))得到的結果，變換後的曲線基本上重合在一起。

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100266

第五章 結論

金融數據中的價差(return)分布呈現厚尾、具冪次定律(power-law)性質，股價上升或下跌的持續時間之分布亦呈現冪次定律[13][14]。本論文以一維隨機模型、將股價的下跌與上漲兩種狀況，分別視為二位元中的 0 與 1，藉以尋求可能產生冪次定律的情況。我們的理論分析與電腦模擬結果顯示，連續 1 的字串長度分布基本上為指數函數，而進一步分析由字串歷史所決定的、有記憶之隨機過程，我們得到:在固定記憶長度的情況，字串會往記憶視窗全為 1 或全為 0 的狀態收斂，而其平均收斂時間隨視窗大小呈現冪次定律的形式遞增;且經由尺度變換後，不同視窗長度下的機率密度分布曲線會重疊在一起。此結果可能暗示著金融數據中的冪定律性質或可經由進一步的時間尺度變換的不變性予以了解。

在社會物理的研究中，代理人模型(agent-based model)，常常探討系統狀態，依代理人之記憶，在若干給定的狀態中做選擇，例如曾嘉瑤的碩士論文中[19]，討論到經由學習做決策，即為一種記憶的形式產生效應；而系統最終也收斂到固定點，或在固定點間做震盪。若將我們的模擬中現有的兩種位元擴充為更多樣的型態，則可將我們的方法應用於各種社會物理議題。

DOI:10.6814/NCCU202100266

附錄 A

參考文獻[16]中定義總長度為n的二進位字串的分布與n個伯努利試驗x_1、x_2、⋯、x_n中，成功機率為p，$0 \le p \le 1$，失敗機率為q，$q=1-p$的性質相關。總長度為n的二位元字串中，定義R_L為連續L個 1 出現的連續列，且任意兩個連續列間必至少間隔一個 0，例如:總長度$n=8$的二進位字串 00111011 $\Rightarrow 00R_3 0R_2$ 。今考慮$1 \le L \le n-1$的範圍下，在2^n個可能的字串(string)逐一檢查長度為L的連續列出現的個數，令其出現的總個數為N_L，其中k為總長度為n的二進位字串扣除一段長度為L的連續列後剩餘的個數，即$k=n-L$，則以下為n、L、k、N_L的關係。

n：二進位字串的總長度

L：連續 1 出現的長度

k：總長度n扣除長度為L的連續列後所剩餘的字串個數

N_L：在2^n個可能的字串中，長度L的連續列出現的總個數

N_L代表發生所有R_L情況的總個數，則有底下的關係式：

$$\begin{cases} N_{n-k} = 2N_{n-k+1} + 2^{k-2}, k \ge 2 \\ N_n = 1 \\ N_{n-1} = 2 \end{cases} \tag{A.1}$$

且得 $N_L = N_{n-k} = (k+3) \cdot 2^{k-2}$, $k \ge 2$ (A.2)

DOI:10.6814/NCCU202100266

在上面的計數方式中，我們首要的工作，即為用簡單的排列組合方式加以解釋，並確認其計數方式中是否有重複的連續列，以字串總長度 $n=8$ 為例，以列舉的方式條列其中的 $1 \le k \le 5$，解釋(A.1)式

(1) $k=1$，$L=7 \Rightarrow N_7=2$

01111111

11111110

(2) $k=2$，$L=6 \Rightarrow N_6=2N_7+1=5$

01111111 ⇒ 00111111、10111111

11111110 ⇒ 11111100、11111101

少 ⇒ 01111110

(3) $k=3$，$L=5 \Rightarrow N_5=2N_6+2=12$

00111111 ⇒ 00011111、01011111

10111111 ⇒ 10011111、11011111

01111110 ⇒ 00111110、01111100

11111101 ⇒ 11111001、11111011

11111100 ⇒ 11111000、11111010

少 ⇒ 10111110、01111101

DOI:10.6814/NCCU202100266

(4) $k=4$, $L=4 \Rightarrow N_4 = 2N_5 + 2^2 = 28$

00011111 ⇒ 00001111、00101111

10011111 ⇒ 10001111、10101111

01011111 ⇒ 01001111、01101111

11011111 ⇒ 11001111、11101111

00111110 ⇒ 00011110、01011110

10111110 ⇒ 10011110、11011110

01111101 ⇒ 01111001、01111011

01111100 ⇒ 01111000、01111010

11111011 ⇒ 11110011、11110111

11111010 ⇒ 11110010、11110110

11111001 ⇒ 11110001、11110101

11111000 ⇒ 11110000、11110100

少 ⇒ 00111100、10111100、00111101、10111101

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100266

(5) $k=5$, $L=3 \Rightarrow N_3 = 2N_4 + 2^3 = 64$

個數為 64 個(其中 11100111、01110111、11101110 各算兩次)

(再加 11110111、11101111)

00001111 ⇒ 00000111、00010111　　00101111 ⇒ 00100111、00110111

10101111 ⇒ 10100111、10110111　　10001111 ⇒ 10000111、10010111

01101111 ⇒ 01100111、$\boxed{01110111}$　　01001111 ⇒ 01000111、01010111

11101111 ⇒ 11100111、11110111　　11001111 ⇒ 11000111、11010111

01011110 ⇒ 01001110、01101110　　00011110 ⇒ 00001110、00101110

10011110 ⇒ 10001110、10101110　　11011110 ⇒ 11011100、11011101

10111100 ⇒ 10011100、10111000

00111100 ⇒ 00011100、00111000

10111101 ⇒ 10011101、10111001

00111101 ⇒ 00111001、00011101

01111001 ⇒ 01110001、01110101　　01111011 ⇒ 00111011、10111011

01111010 ⇒ 01110010、01110110　　01111000 ⇒ 01110000、01110100

11110111 ⇒ 11100111、11101111　　11110011 ⇒ 11100011、11101011

11110110 ⇒ 11100110、$\boxed{11101110}$　　11110010 ⇒ 11100010、11101010

11110101 ⇒ 11100101、11101101　　11110001 ⇒ 11100001、11101001

11110000 ⇒ 11100000、11101000　　11110100 ⇒ 11100100、11101100

少 ⇒ 01110011、11001110、10111010、01011101、00111010、01011100

DOI:10.6814/NCCU202100266

上述的條列過程中，我們得到下述兩種性質

① $n-(L+1)\geq L$ 時有兩個以上長度為 L

② n 字串中，最多有 $\left[\dfrac{n+1}{L+1}\right]$ 個 $\underbrace{111\cdots11}_{L}$

在此我們證明第②個性質：

設 n 字串中最多有 x 個 $\underbrace{111\cdots11}_{L}$，則至少有 $x-1$ 個 0

$$\Rightarrow L\cdot x+(x-1)\leq n \Rightarrow (n-k+1)\cdot x\leq n+1 \Rightarrow x\leq \frac{n+1}{n-k+1}$$

即 n 字串中最多有 $\left[\dfrac{n+1}{L+1}\right]$ 個 $\underbrace{111\cdots11}_{L}$

而關於(A.2)式的關係式，我們以簡單的排列組合計數方式在所有 2^n 個可能的二位元字串中，計算長度為 L 的總出現次數 N_L：

只考慮計算連續 L 個 1 的總出現次數，可視為 $\underbrace{111\cdots111}_{L}$，$\underbrace{*,*\cdots,*}_{k}$ 的直線排列，

(1) ($\underbrace{111\cdots111}_{L}0$，$\underbrace{*,*\cdots,*}_{k-1}$) 或 ($\underbrace{*,*\cdots,*}_{k-1}$，$0\underbrace{111\cdots111}_{L}$) 共 $2^{k-1}\times 2$ 個可能

(2) $0\underbrace{111\cdots111}_{L}0$，$\underbrace{*,*\cdots,*}_{k-2}$ 共 $C_{k-2}^{k-1}\times 2^{k-2}$ 個可能

$$\Rightarrow 2^{k-1}\times 2+C_{k-2}^{k-1}\times 2^{k-2}=(k+3)\times 2^{k-2}$$

即 $N_L=N_{n-k}=(k+3)\cdot 2^{k-2}$，$k\geq 2$

DOI:10.6814/NCCU202100266

附錄 B

在本文 3.2 節中計算機率質量函數 $P(L)$ 的過程如下，首先我們先建立下列四個公式:

級數 1：$\sum_{k=1}^{a} k^2 \cdot 2^{k-\ell} = (a^2 - 2a + 3) \cdot 2^{a-\ell+1} - 3 \cdot 2^{1-\ell}$ （B1）

級數 2：$\sum_{k=1}^{a} k \cdot 2^{k-\ell} = (a-1) \cdot 2^{a-\ell+1} + 2^{1-\ell}$ （B2）

級數 3：$\sum_{k=1}^{a} 2^{k-\ell} = 2^{a-\ell+1} - 2^{1-\ell}$ （B3）

級數 4：$\sum_{k=1}^{a} k^3 \cdot 2^{k-\ell} = (a^3 - 3a^2 + 9a - 13) \cdot 2^{a-\ell+1} + 13 \cdot 2^{1-\ell}$ （B4）

推導過程為：

(B1)：$\sum_{k=1}^{a} k^2 \cdot 2^{k-\ell} = (a^2 - 2a + 3) \cdot 2^{a-\ell+1} - 3 \cdot 2^{1-\ell}$

$<pf>$：令 $S_1 = \sum_{k=1}^{a} k^2 \cdot 2^{k-\ell} = 1^2 \cdot 2^{1-\ell} + 2^2 \cdot 2^{2-\ell} + 3^2 \cdot 2^{3-\ell} + \cdots + a^2 \cdot 2^{a-\ell}$

$$
\begin{aligned}
S_1 &= 1^2 \cdot 2^{1-\ell} + 2^2 \cdot 2^{2-\ell} + \cdots + \quad a^2 \cdot 2^{a-\ell} \\
2S_1 &= \qquad\quad + 1^2 \cdot 2^{2-\ell} + \cdots + (a-1)^2 \cdot 2^{a-\ell} + a^2 \cdot 2^{a-\ell+1} \quad (- \\
\hline
-S_1 &= 1 \cdot 2^{1-\ell} + 3 \cdot 2^{2-\ell} + \cdots + (2a-1) \cdot 2^{a-\ell} - a^2 \cdot 2^{a-\ell+1} \qquad \cdots\cdots\cdots(1)
\end{aligned}
$$

令 $S_1' = 1 \cdot 2^{1-\ell} + 3 \cdot 2^{2-\ell} + 5 \cdot 2^{3-\ell} + \cdots + (2a-1) \cdot 2^{a-\ell}$

$$
\begin{aligned}
S_1' &= 1 \cdot 2^{1-\ell} + 3 \cdot 2^{2-\ell} + \cdots + (2a-1) \cdot 2^{a-\ell} \\
2S_1' &= \qquad\quad + 1 \cdot 2^{2-\ell} + \cdots + (2a-3) \cdot 2^{a-\ell} + (2a-1) \cdot 2^{a-\ell+1} \quad (- \\
\hline
-S_1' &= 1 \cdot 2^{1-\ell} + 2 \cdot 2^{2-\ell} + \cdots + \quad 2 \quad \cdot 2^{a-\ell} - (2a-1) \cdot 2^{a-\ell+1} \\
&= 2^{1-\ell} + 2 \cdot 2^{2-\ell} \cdot (2^{a-1} - 1) - (2a-1) \cdot 2^{a-\ell+1} \\
&= (-2a+3) \cdot 2^{a-\ell+1} - 3 \cdot 2^{1-\ell} \\
\Rightarrow S_1' &= (2a-3) \cdot 2^{a-\ell+1} + 3 \cdot 2^{1-\ell} \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(2)
\end{aligned}
$$

DOI:10.6814/NCCU202100266

將(2)式代回(1)式得

$$-S_1=(2a-3)\cdot 2^{a-\ell+1}+3\cdot 2^{1-\ell}-a^2\cdot 2^{a-\ell+1}=(-a^2+2a-3)\cdot 2^{a-\ell+1}+3\cdot 2^{1-\ell}$$

則 $S_1=\sum_{k=1}^{a}k^2\cdot 2^{k-\ell}=(a^2-2a+3)\cdot 2^{a-\ell+1}-3\cdot 2^{1-\ell}$

(B2)： $\sum_{k=1}^{a}k\cdot 2^{k-\ell}=(a-1)\cdot 2^{a-\ell+1}+2^{1-\ell}$

$<pf>$：令 $S_2=\sum_{k=1}^{a}k\cdot 2^{k-\ell}=1\cdot 2^{1-\ell}+2\cdot 2^{2-\ell}+3\cdot 2^{3-\ell}+\cdots+a\cdot 2^{a-\ell}$

$$\begin{aligned}
S_2 &= 1\cdot 2^{1-\ell}+2\cdot 2^{2-\ell}+3\cdot 2^{3-\ell}+\cdots+\quad a\cdot 2^{a-\ell}\\
2S_2 &= \qquad\quad +1\cdot 2^{2-\ell}+2\cdot 2^{3-\ell}+\cdots+(a-1)\cdot 2^{a-\ell}+a\cdot 2^{a-\ell+1} \quad (-\\
\hline
-S_1 &= \quad 2^{1-\ell}+\ 2^{2-\ell}+\ 2^{3-\ell}+\cdots\cdots\cdots+\ 2^{a-\ell}\ -\ a\cdot 2^{a-\ell+1}\\
&= 2^{1-\ell}\cdot(2^a-1)-a\cdot 2^{a-\ell+1}\\
&= (1-a)\cdot 2^{a-\ell+1}-2^{1-\ell}
\end{aligned}$$

則 $S_2=\sum_{k=1}^{a}k\cdot 2^{k-\ell}=(a-1)\cdot 2^{a-\ell+1}-2^{1-\ell}$

(B3)： $\sum_{k=1}^{a}2^{k-\ell}=2^{a-\ell+1}-2^{1-\ell}$

$<pf>$： $\sum_{k=1}^{a}2^{k-\ell}=2^{1-\ell}+2^{2-\ell}+2^{3-\ell}+\cdots+2^{a-\ell}=2^{1-\ell}(2^a-1)=2^{a-\ell+1}-2^{1-\ell}$

(B4)： $\sum_{k=1}^{a}k^3\cdot 2^{k-\ell}=(a^3-3a^2+9a-13)\cdot 2^{a-\ell+1}+13\cdot 2^{1-\ell}$

$<pf>$：令 $S_3=\sum_{k=1}^{a}k^3\cdot 2^{k-\ell}=1^3\cdot 2^{1-\ell}+2^3\cdot 2^{2-\ell}+3^3\cdot 2^{3-\ell}+\cdots+a^3\cdot 2^{a-\ell}$

其中 $(n+1)^3-n^3=3n^2+3n+1$

DOI:10.6814/NCCU202100266

$$S_3 = 1^3\cdot 2^{1-\ell} + \quad 2^3\cdot 2^{2-\ell} \quad +\cdots+ \quad a^3\cdot 2^{a-\ell}$$

$$2S_3 = \qquad\qquad + \quad 1^3\cdot 2^{2-\ell} \quad +\cdots+(a-1)^3\cdot 2^{a-\ell}+a^3\cdot 2^{a-\ell+1} \quad (-$$

$$-S_3 = 1\cdot 2^{1-\ell}+(3\cdot 1^2+3\cdot 1+1)\cdot 2^{2-\ell}+\cdots+(3\cdot(a-1)^2+3\cdot(a-1)+1)\cdot 2^{a-\ell}-a^3\cdot 2^{a-\ell+1}$$

$$=1\cdot 2^{1-\ell}+3\sum_{i=1}^{a-1} i^2\cdot 2^{i+1-\ell}+3\sum_{i=1}^{a-1} i\cdot 2^{i+1-\ell}+\sum_{i=1}^{a-1} 2^{i+1-\ell}-a^3\cdot 2^{a-\ell+1}$$

$$=1\cdot 2^{1-\ell}+6\sum_{i=1}^{a-1} i^2\cdot 2^{i-\ell}+6\sum_{i=1}^{a-1} i\cdot 2^{i-\ell}+2\sum_{i=1}^{a-1} 2^{i-\ell}-a^3\cdot 2^{a-\ell+1}$$

代入(B1)、(B2)、(B3)得

$$-S_3 = (-a^3+3a^2-9a+13)\cdot 2^{a-\ell+1}-13\cdot 2^{1-\ell}$$

則 $S_3 = \sum_{k=1}^{a} k^3\cdot 2^{k-\ell} = (a^3-3a^2+9a-13)\cdot 2^{a-\ell+1}+13\cdot 2^{1-\ell}$

$$=\frac{(n^2-n-4)\cdot 2^{2n-3}-(n^3-3n^2+4n-10)\cdot 2^{n-2}-2}{[(n+1)\cdot 2^{n-2}-1]^2}$$

機率質量函數為 $P(L)=\dfrac{(n-L+3)\cdot 2^{n-L-2}}{(n+1)\cdot 2^{n-2}-1}$ ， $1\le L\le n-1$ 的公式證明如下:

$<pf>$：附錄 B 中，(B2)、(B3)的式子令 $a=n-1$ 、 $\ell=2$ 代入

$$\sum_{k=1}^{n-1}(k+3)\cdot 2^{k-2}=\sum_{k=1}^{n-1} k\cdot 2^{k-2}+3\sum_{k=1}^{n-1} 2^{k-2}$$

$$=[(n-2)\cdot 2^{n-2}+2^{-1}]+3\cdot[2^{n-2}-2^{-1}]$$

$$=(n+1)\cdot 2^{n-2}-1$$

$$\Rightarrow P(L)=P(n-k)$$

$$=\frac{(k+3)\cdot 2^{k-2}}{\sum_{k=1}^{n-1}(k+3)\cdot 2^{k-2}}$$

$$=\frac{(k+3)\cdot 2^{k-2}}{(n+1)\cdot 2^{n-2}-1}$$

$$=\frac{(n-L+3)\cdot 2^{n-L-2}}{(n+1)\cdot 2^{n-2}-1} \quad , \quad 1\le L\le n-1$$

其中： $\sum_{L=1}^{n-1} P(L)=\sum_{L=1}^{n-1}\dfrac{(k+3)\cdot 2^{k-2}}{(n+1)\cdot 2^{n-2}-1}=1$

DOI:10.6814/NCCU202100266

期望值 $E(L)=\dfrac{n\cdot(2^{n-1}-1)}{(n+1)\cdot 2^{n-2}-1}$ ， $1\le L\le n-1$ 的證明如下:

$<pf>$：附錄 B 中，(B1)、(B2)、(B3)的式子中令 $a=n-1$ 、 $\ell=2$ 代入

$$\begin{aligned}
&\sum_{k=1}^{n-1}(k+3)\cdot 2^{k-2}\cdot(n-k)\\
&=-\sum_{k=1}^{n-1}k^2\cdot 2^{k-2}+(n-3)\cdot\sum_{k=1}^{n-1}k\cdot 2^{k-2}+3n\cdot\sum_{k=1}^{n-1}2^{k-2}\\
&=n\cdot(2^{n-1}-1)
\end{aligned}$$

$$\begin{aligned}
\Rightarrow E(L)&=E(n-k)\\
&=\sum_{k=1}^{n-1}P(n-k)\cdot(n-k)\\
&=\sum_{k=1}^{n-1}\frac{(k+3)\cdot 2^{k-2}\cdot(n-k)}{(n+1)\cdot 2^{n-2}-1}\\
&=\frac{n\cdot(2^{n-1}-1)}{(n+1)\cdot 2^{n-2}-1}
\end{aligned}$$

變異數 $Var(L)=\dfrac{(n^2-n-4)\cdot 2^{2n-3}-(n^3-3n^2+4n-10)\cdot 2^{n-2}-2}{[(n+1)\cdot 2^{n-2}-1]^2}$ ， $1\le L\le n-1$ 的證明如下:

$<pf>$：其中

$$\begin{aligned}
E(L^2)&=E[(n-k)^2]\\
&=\sum_{k=1}^{n-1}P(n-k)\cdot(n-k)^2\\
&=\sum_{k=1}^{n-1}\frac{(k+3)\cdot 2^{k-2}\cdot(n-k)^2}{(n+1)\cdot 2^{n-2}-1}\\
&=\frac{(3n-4)\cdot 2^{n-1}-n^2+2}{(n+1)\cdot 2^{n-2}-1}
\end{aligned}$$

則

$$\begin{aligned}
Var(L)&=Var(n-k)\\
&=E[(n-k)^2]-[E(n-k)]^2\\
&=\frac{(n^2-n-4)\cdot 2^{2n-3}-(n^3-3n^2+4n-10)\cdot 2^{n-2}-2}{[(n+1)\cdot 2^{n-2}-1]^2}
\end{aligned}$$

DOI:10.6814/NCCU202100266

附錄 C

在 3.2.2 節(3.4)式中，持續次數$\tau(L,s)$的機率質量函數$P(\tau)$的證明如下:

$<pf>$：在長度為L的連續列中，長度為s的次連續列以重疊計數的方式計算共出現了$\tau = L - s + 1$次，則於所有長度為$L \geq s$的連續列中，持續次數τ的機率質量函數

$$
\begin{aligned}
P(\tau) &= P(L-s+1) \\
&= P(n-k-s+1) \\
&= \frac{(k+3)\cdot 2^{k-2}\cdot(n-k-s+1)}{\sum_{k=1}^{n-s}(k+3)\cdot 2^{k-2}\cdot(n-k-s+1)} \\
&= \frac{(k+3)\cdot 2^{k-2}\cdot(n-k-s+1)}{(n-s+1)[2^{n-s}-1]} \\
&= \frac{(n-\tau-s+4)\cdot 2^{n-\tau-s-1}\cdot\tau}{(n-s+1)[2^{n-s}-1]}
\end{aligned}
$$

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100266

附錄 D

在 4.3.2 節，轉移矩陣 $\boldsymbol{A}$((4.1)式)所描述式窗狀態的變動，其中的固定點狀態可由下列等式獲得:

$$\begin{pmatrix} 1 & \frac{m-1}{m^2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{(m-1)^2+1}{m^2} & \frac{2(m-2)}{m^2} & & & 0 & 0 \\ 0 & \frac{m-1}{m^2} & \frac{(m-2)^2+4}{m^2} & & & & 0 \\ \vdots & 0 & \frac{2(m-2)}{m^2} & & & & \vdots \\ \vdots & & 0 & \ddots & 0 & & \vdots \\ \vdots & & & & \frac{2(m-2)}{m^2} & 0 & \vdots \\ 0 & & & & \frac{(m-2)^2+4}{m^2} & \frac{m-1}{m^2} & 0 \\ 0 & 0 & & & \frac{2(m-2)}{m^2} & \frac{(m-1)^2+1}{m^2} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{m-1}{m^2} & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ \vdots \\ a_{m-2} \\ a_{m-1} \\ a_m \end{pmatrix} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ \vdots \\ a_{m-2} \\ a_{m-1} \\ a_m \end{pmatrix}$$

可解出 $a_0 + \frac{m-1}{m^2} a_1 = a_0 \Rightarrow a_1 = 0$，以及 $a_2 = a_3 = \cdots\cdots = a_{m-1} = 0$，此處 $a_1 + a_2 + \cdots\cdots + a_m = 1$，所描述的狀態為視窗中有 $m-j$ 個 1、j 個 0 的情況有 a_j 的機率，滿足 $a_0 + a_m = 1$ 的選擇都是解。其中 $a_0 = 1$、$a_m = 0$ 及 $a_0 = 0$、$a_m = 1$ 分別為單純的全為 1 及全為 0 的狀態。

DOI:10.6814/NCCU202100266

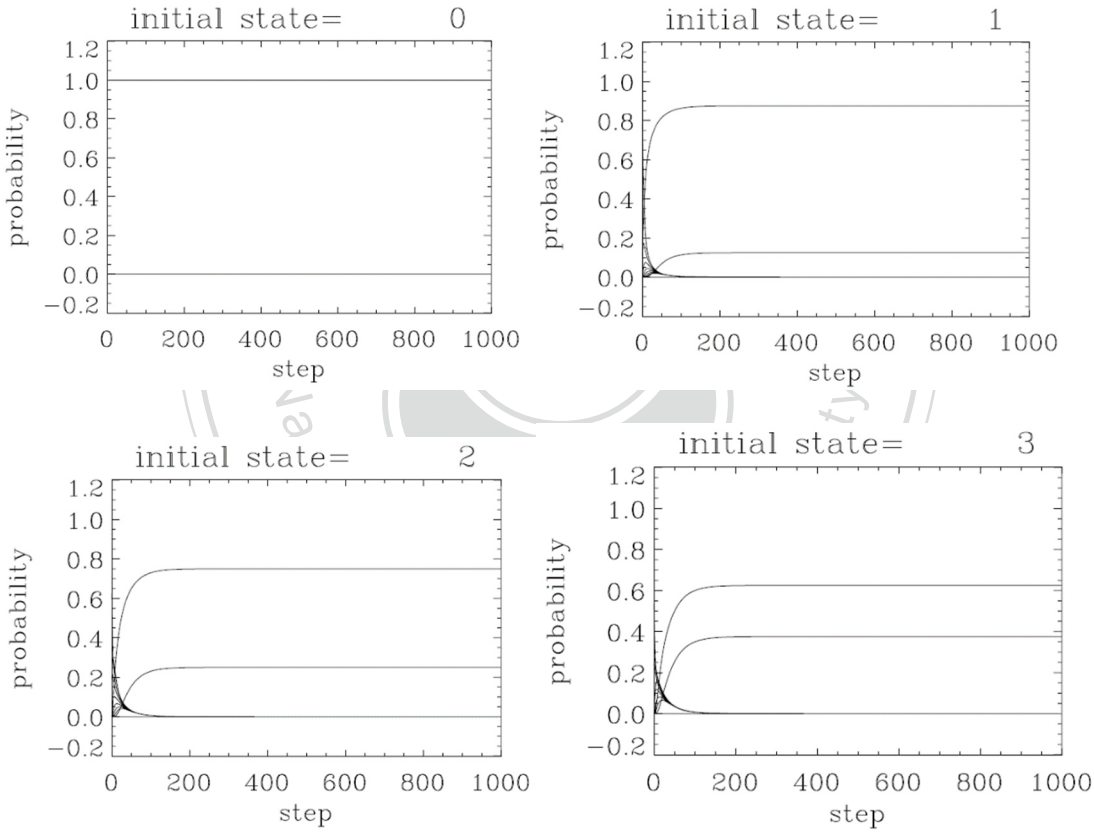
附錄 E

考慮是記憶長度 $m=8$，由 $\boldsymbol{A}$ 矩陣(式(4.1))的 M 次方乘積，計算得到初始狀態為

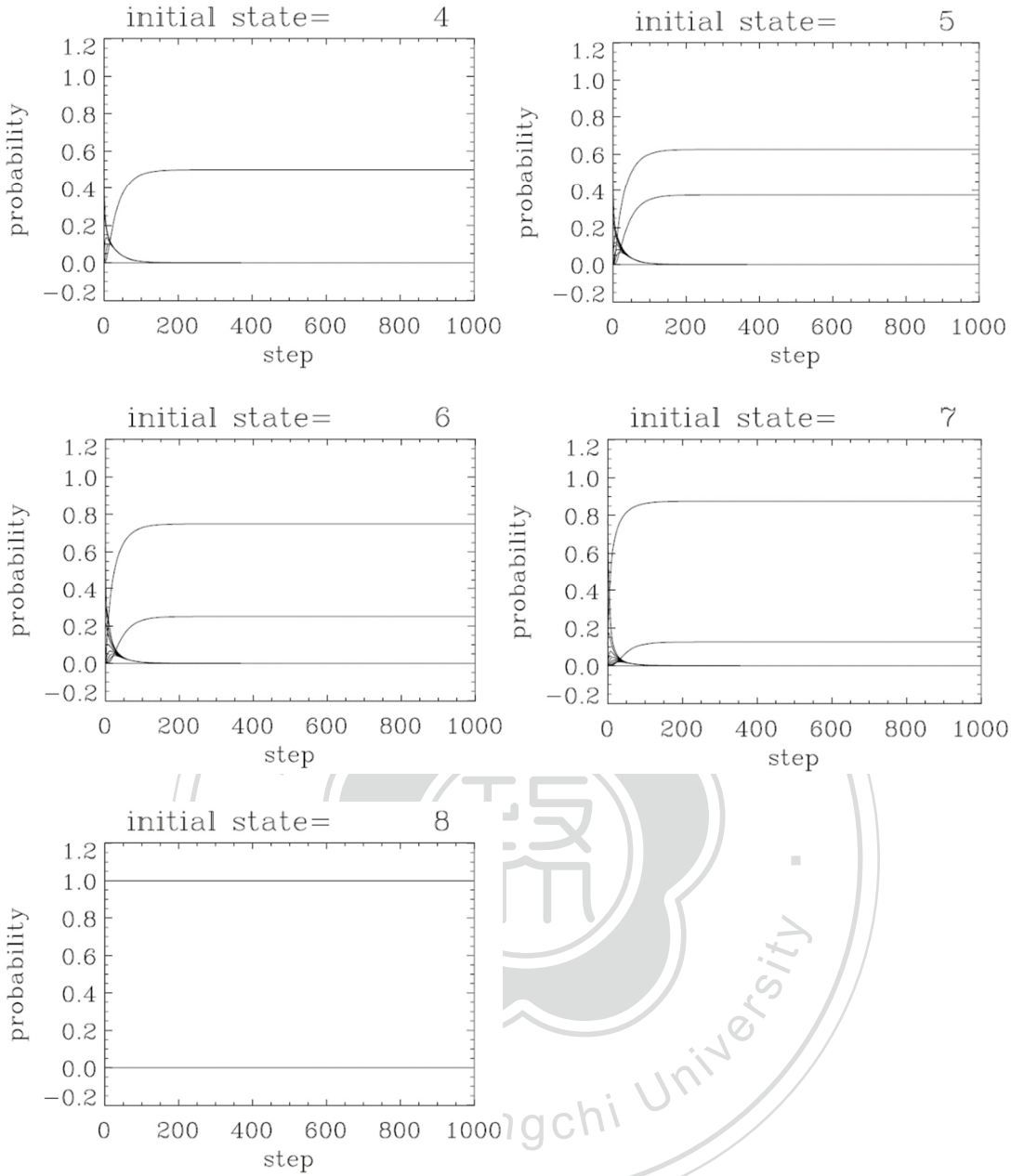
(圖中初始狀態 j) $a_j=1$，其餘 $a_k=0\,(k\neq j)$，$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix}$ 隨步數 M 的變動。可得到：

由初始狀態 j 出發，最後收斂為 $a_0=\dfrac{m-j}{m}$，$a_m=\dfrac{j}{m}$，其他

$a_k=0\,(k\neq 0, k\neq m)$。



DOI:10.6814/NCCU202100266



DOI:10.6814/NCCU202100266

參考文獻

[1] E. Dimson, M. Mussavian. " A brief history of market efficiency", *European Financial Management* 4 (1998): 91-193.

[2] J. Voit. *The statistical mechanics of financial markets.* (Springer, 2013).

[3] M. F. Osborne. "Brownian motion in the stock market." *Operations research* 7 (1959):145-173.

[4] E. F. Fama. "The behavior of stock-market prices." *The journal of Business* 38 (1965):34-105.

[5] R. N. Mantegna, H. E. Stanley. "Scaling behavior in the dynamics of an economic index." *Nature* 376 (1995):46-49.

[6] R. Cont. "Empirical properties of asset returns: stylized facts and statistical issues." (2001):223-236.

[7] V. Plerou *et al*. "Universal and nonuniversal properties of cross correlations in financial time series." *Phys. Rev. Lett.* 83 (1999) :1471.

[8] L. Laloux, P. Cizeau, J. P. Bouchaud, and M. Potters. "Random matrix theory and financial correlations." *Phys. Rev. Lett.* 83 (1999) :1467.

[9] W. J. Ma, C. K. Hu, and R. E. Amritkar. "Stochastic dynamical model for stock-stock correlations." *Physical Review E* 70 (2004):026101.

[10] W. J. Ma, et al. "Crossover behavior of stock returns and mean square displacements of particles governed by the Langevin equation. " *EPL* 102 (2013):66003.

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100266

[11] V. Kishore, M. S. Santhanam, R. E. Amritkar."Extreme events and event size fluctuations in biased random walks on networks." *Physical Review E* 85 (2012):056120.

[12] C. W. Chen, W. J. Ma."Toward ascenario with complementary stochastic and deterministic information in financial fluctuations." *Chinese Journal of Physics* 56 (2018):853-862.

[13] 施奕甫，「股票市場中事件發生的尺度性質」碩士論文，國立政治大學應用物理研究所(2017).

[14] 吳培煜，「股票變化之波動的若干流體動力訊號－跨天效應」碩士論文，國立政治大學應用物理研究所(2018).

[15] N. Balakrishnan, and Markos V. Koutras. *Runs and scans with applications.* (John Wiley & Sons, 2011).

[16] F. S. Makri, Z. M. Psillakis."On success runs of a fixed length in Bernoulli sequences: Exact and asymptotic results." *Computers & Mathematics with Applications* 61 (2011):761-722.

[17] K. Sinha, B. P. Sinha."On the distribution of runs of ones in binary strings." *Computers & Mathematics with Applications* 58 (2009):1816-1829.

[18] 黃文璋, *數理統計*, 華泰 , 2003.

[19] 曾嘉瑤，「以代理人基模型模擬的施與受賽局」碩士論文，國立政治大學應用物理研究所(2013).

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100266