

國立交通大學

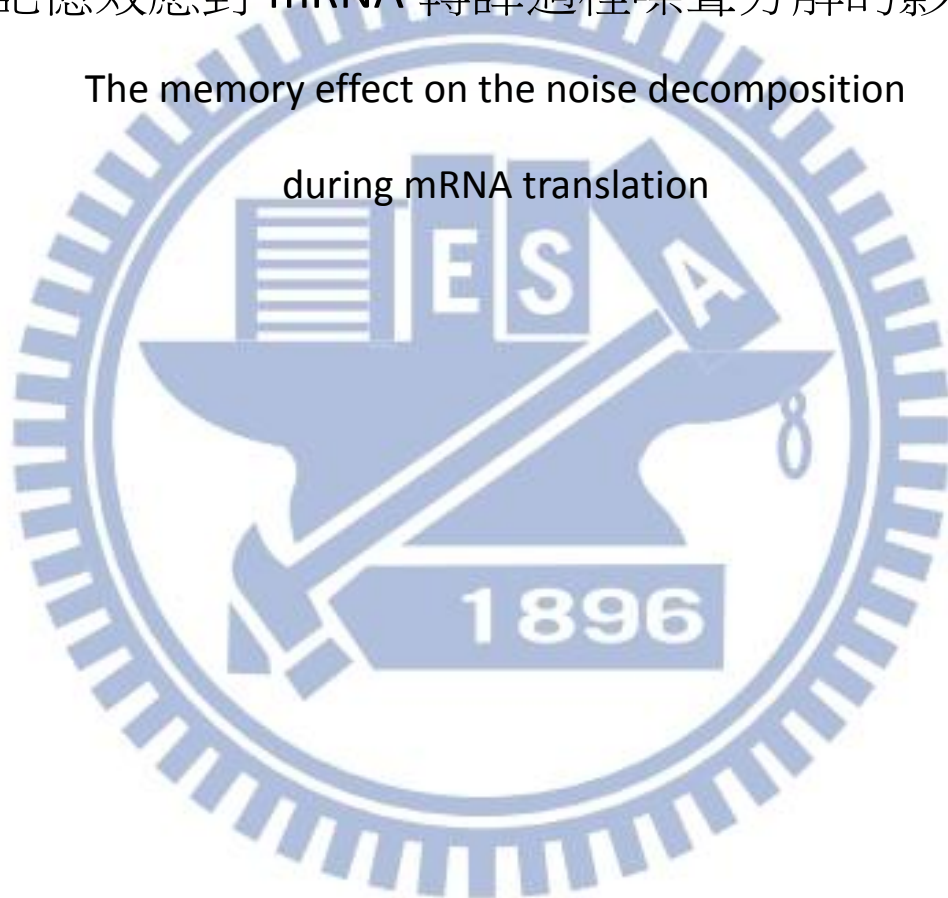
物理研究所

碩士論文

記憶效應對 mRNA 轉譯過程噪聲分解的影響

The memory effect on the noise decomposition

during mRNA translation



研究生：黃冠智

指導教授：張正宏 副教授

中華民國一百零三年七月

記憶效應對 mRNA 轉譯過程噪聲分解的影響

The memory effect on the noise decomposition
during mRNA translation

研究生:黃冠智

Student: Guan-Jhih, Huang

指導教授:張正宏

Advisor: Cheng-Hung Chang



國立交通大學

物理研究所

碩士論文

A Thesis

Submitted to Institute of Physics College of Science

National Chiao Tung University

In partial Fulfillment of the Requirements

For the degree of

Master

In

Physics

July 2014

Hsinchu, Taiwan, Republic of China

中國民國一零四年七月

記憶效應對 mRNA 轉譯過程噪聲分解的影響

研究生:黃冠智

指導教授:張正宏

國立交通大學物理研究所碩士班

中文摘要

在生活中處處存在著噪聲(noise)，而如何分析這些噪聲的來源是相當重要的。我們使用 dual-reporters 的方法將噪聲拆解成內部與外部噪聲，並分別以當下環境的狀態和環境的歷史為條件將 mRNA 轉譯成 protein 過程中的噪聲拆解，並利用 variance norm 與 KL_divergence 來了解當下的 protein 與此不同時間之前 mRNA 的關聯。

由上述方法得知，在大部分的情況下，只考慮當下環境的狀態為條件做拆解時，並不能正確的拆解噪聲，而必須考慮所有環境的歷史才能精確的拆解內外噪聲，不過在特殊的狀況下：如在 mRNA 的耗散速率遠大於 protein 的耗散速率時，以當下環境的狀態就能得到相當好的拆解，且環境歷史的影響程度會因為在不同的速率常數下各不相同。

關鍵詞:擾動、噪音、分解

The memory effect on the noise decomposition during mRNA translation

Student: Guan-Jhih, Huang

Advisor: Cheng-Hung Chang

Institute of Physics

National Chiao Tung University

ABSTRACT

Noises are everywhere in our life. How to analyze the sources of these noises is very important. We use the dual-reporters method to separate noises into extrinsic and intrinsic parts. We also separate the noises in mRNA translation, by using the present state or other historical states of the environment. Then we analyze the relation between the present status of protein and different lengths of history of mRNA by using the variance norm and KL divergence. In most of the situations, we cannot separate the noises correctly by using the condition on the states of the present environment, We need to consider the full histories of the environment to separate the noises correctly. However, in some particular cases, like when the dissipation rate of mRNA is greater than that of protein, we can separate the noises correctly by the present state of the environment. The effects of the histories of environment are different in different rate constants.

Keywords : Fluctuation, Noise

誌謝

首先感謝家人讓我有這個機會到交通大學就讀物理研究所，有了家中的支援協助才能專心的在這裡學習，並多次的北上關心生活起居與參與畢業典禮。

也非常感謝指導教授，張正宏副教授，在碩一下能收我為指導學生，並在碩士生活中學到了許多物裡的知識與研究的精神，也在每周的討論中學習到了教授如何清楚的表達自己的意見，做研究的方法與嚴謹的態度，讓研究生活有個充實的一面。

另外感謝研究室的德明跟弘慶學長，不吝嗇的與我們分享經驗與討論問題。研究室的夥伴，賜得與禮凡，在研究生的生活中同伴是不可或缺的，謝謝有你們常常一起討論問題一起分享生活，讓研究生活更多了股活力。

最後感謝在交大認識的依璇，在這一年中的陪伴與鼓勵，碩班生活可為多彩多姿，謝謝一路上的每個人。



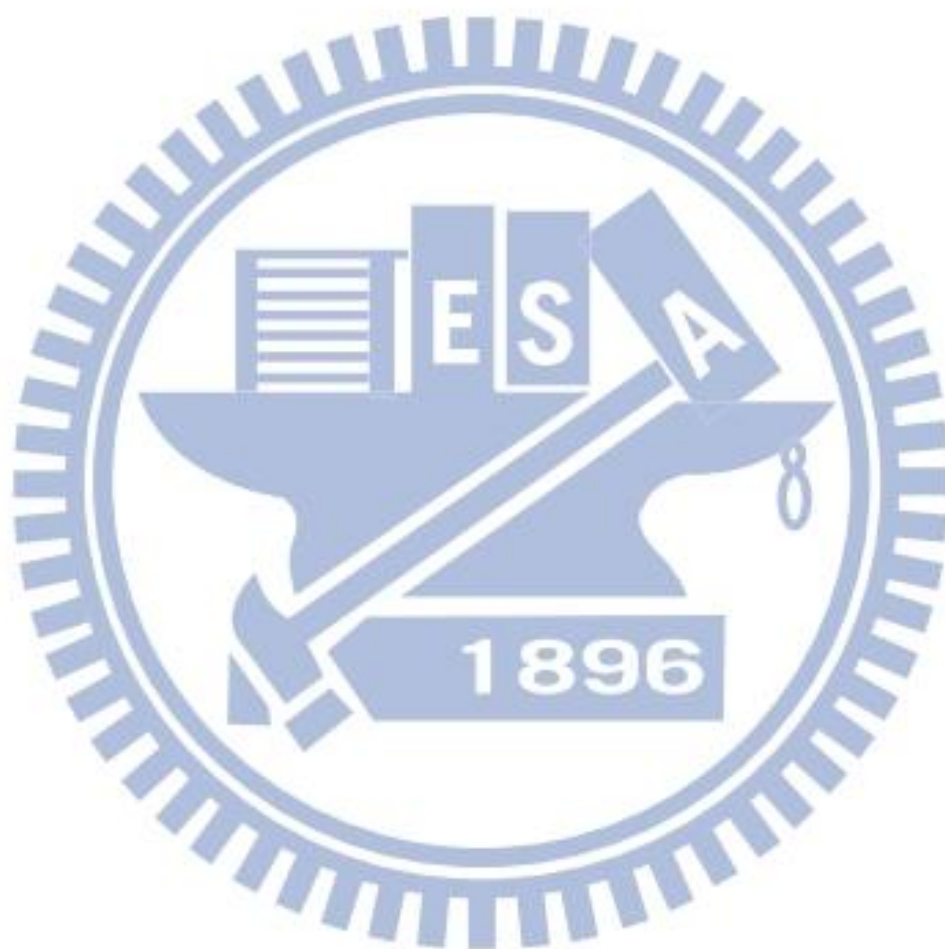
目錄

中文摘要.....	i
英文摘要.....	ii
誌謝.....	iii
目錄.....	iv
表列.....	v
第 1 章 緒論	1
第 2 章 理論	3
2.1 建立將擾動區分為內部及外部噪音的公式	3
2.2 利用 Dual-Reporter Method 來區分訊號.....	7
2.3 利用當下環境的狀態來分解噪音	7
2.4 利用環境的歷史來分解噪音	10
2.5 KL_Divergence.....	12
第 3 章 系統架構	13
3.1 系統	13
3.2 數值模擬方法	14
第 4 章 研究結果	15
4.1 利用當下環境的狀態來分解噪音	15
4.2 利用環境的歷史來分解噪音	21
4.3 X 與不同時間之前的 Z 的關聯.....	27
4.4 當環境變動很劇烈時的 Extrinsic and Intrinsic.....	30
4.5 Time-Correlation	31
第 5 章 結論與未來工作	36
參考文獻.....	37

圖表列

圖 2-1	4
圖 2-2	9
圖 2-3	12
圖 3-1	13
圖 4-1	16
圖 4-2	17
圖 4-3	18
圖 4-4	18
圖 4-5	19
圖 4-6	19
圖 4-7	20
圖 4-8	21
圖 4-9	22
圖 4-10	23
圖 4-11	24
圖 4-12	25
圖 4-13	25
圖 4-14	26
圖 4-15	26
圖 4-16	27
圖 4-17	28
圖 4-18	29
圖 4-19	30

圖 4-20	31
圖 4-21	31
圖 4-22	32
圖 4-23	33
圖 4-24	33
圖 4-25	34



第1章 緒論

在大自然中擾動(fluctuation)的現象相當常見，從細胞中的分子到生態系統中的生物族群都會發生，這些擾動常來自於內在個別事件的隨機性與外在環境的改變。我們希望能夠分析這些擾動因素來源，但是它們之間常常也會互相影響，導致太過複雜而無法做有效的分解，所以之前的研究常會使用簡化的假設，例如忽略某些內部或外部的影響來減少模型自由度的數量。

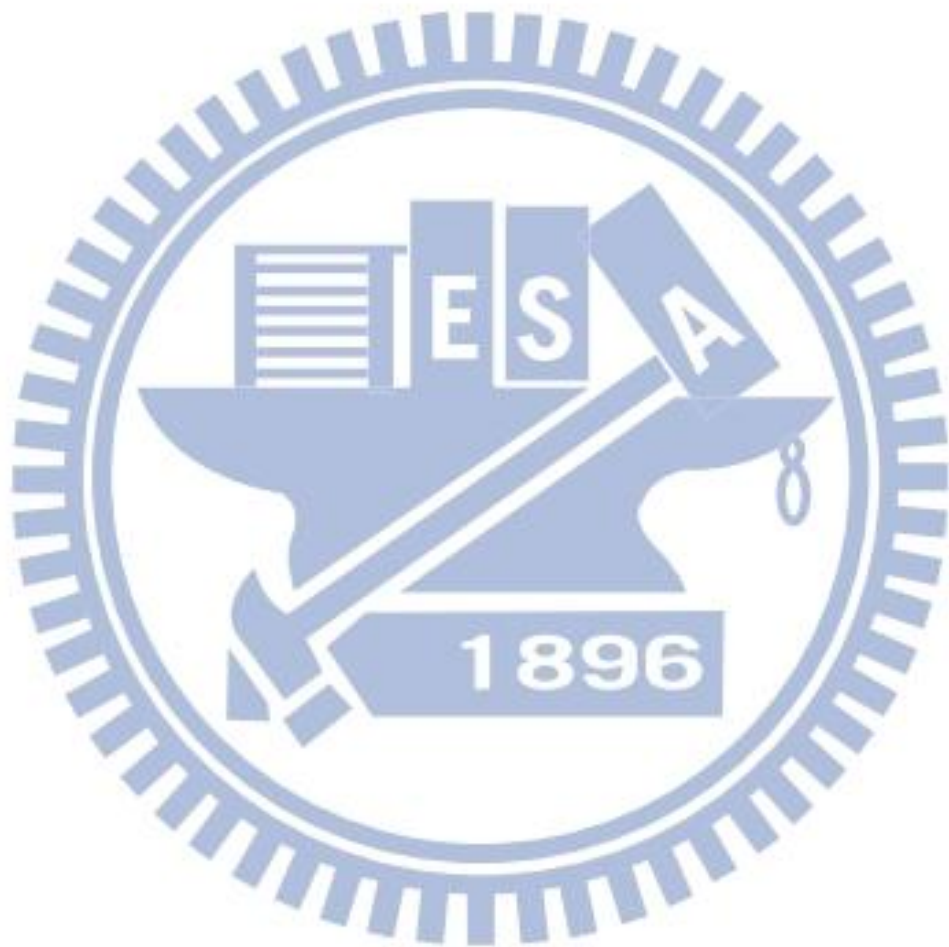
這些擾動如此複雜、如此難以分析，主要是因為這些擾動通常是來自於低數量生死機率(births and deaths at low numbers)的內部噪音與隨機的生物與物理環境的影響。因此，任何給定物種的隨機力學反應了所有直接和間接隨機肇因的連結，這也就是當我們想了解任何給定物種擾動的來源時，會相當複雜且難以分析的原因。

想解決這個看似複雜的問題可以用一個簡單巧妙的方法來解決：在同一個環境中嵌入第二個獨立系統並同時觀察兩個系統。我們可以從系統之間的相關性來了解環境對兩系統的共同影響，而他們之間的差異則反應了每個系統內部產生的自發性噪音。使用這種方法讓我們可以不用知道系統或環境微觀細節的資訊就能區分內部與外部的擾動來源。

實現這種技巧的一種方法是使用 dual-reporter(系統)來識別在 proteins 擾動中細胞層級(即環境)的來源，其中兩 dual-reporter 之間 normalized covariance 使用在基因調控網路內做為外部噪音的測量，且定義其餘的為內部噪音。

為了確保如此分解確實對應到我們平常理解的內外噪聲定義，需要額外做更深入探討。用早期的理論分析，結果說明在共同環境底下總噪聲確實可以利用 dual-reporters 方法分解成平常理解的內部和外部噪音。但這種早期理論的分解只適用於進行靜態環境分析[3]。雖然它在某些特別系統下是對的，但大部分生物系統都是動態的，因此無法使

用。所以在這裡我們想要了解如何在靜態與動態環境的系統中分解內部與外部的擾動、了解他們在生物學上代表著甚麼意思、並嘗試運用它在實際生物模型。

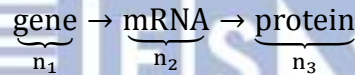


第2章 理論

2.1 建立將擾動區分為內部及外部噪音的公式

基因表達是非常重要且複雜的現象，所以在理論上我們希望去濃縮反應步驟並簡化它，以找出其中的關鍵步驟，即便目前還沒有統一的理論，但在實驗中已經知道類似的基因 [1] 在確切的條件下會使基因產生非常不同的擾動，雖然這讓我們無法確定這些簡化是否適用在特定的高維系統，不過至少可以先透過簡化的模型來了解某些現象。

在基因表達領域，大部分的理論都在描述下面現象：



由上式可以知道 *protein* 的濃度除了會被 *mRNA* 的濃度所影響，間接也會被 *gene* 的濃度影響，其實類似的影響可以無限的向前推，也就是任何間接影響基因表現速度的過程，都有可能影響到 *protein* 濃度而不得被納入模型，但我們在大部分的模型中都只專注在 *genes*、*mRNAs* 及 *proteins* 並將其他所有的過程包含在有效的速度常數中。

其中基因的激活(Gene activation)可以有許多不同的分子原因，包括阻遏物的離解(dissociation of repressors)，活化的關聯(association of activators)或染色質的重建(chromatin remodeling)。但在這裡我們將基因的激活簡化並以下式表達：



其中每個基因自發性的以 λ_1^+ 、 λ_1^- 的速率切換。

在這裡我們忽略了細胞生長和基因複製的影響，在同一段 DNA 中常有數段完全相同的基因，所以我們假設一個常數 n_1^{\max} 在(2.1.1)式中複製獨立的開啟與關閉(假設基因總數為

n_1^{\max})，active genes 數量的機率分布就像是作為丟擲 n_1^{\max} 個不公平硬幣的二項式分布，

$$\text{其中 } P_{\text{on}} = \frac{\lambda_1^+}{\lambda_1^+ + \lambda_1^-}。$$

gene 轉錄成 mRNA 和 mRNA 轉譯成 protein 通常假定遵守 Poisson processes，其中每單位時間內 $\lambda_2 n_1$ 和 $\lambda_3 n_2$ 的生產機率分別正比於 active genes 和 mRNA 的數量 n_1 及 n_2 。

除了上述過程之外，尚有許多不同的因素可能會影響這個動態的系統，包括在許多酵素與基質參與的擾動(fluctuations in the many enzymes and substrates involved)、細胞週期的影響(cell cycle effects)和在細胞分裂中複製的隨機劃分(random partitioning of copies at cell division)等。但標準模型則忽略了這些以及其他複雜的影響，並只關心六個指數級事件(exponential events)：個別基因開啟和關閉的速度常數、active gene的轉錄常數、transcript的轉換常數、transcripts和proteins的指數衰減。

上述介紹總結在下面的反應示意圖和對應的動態行為：

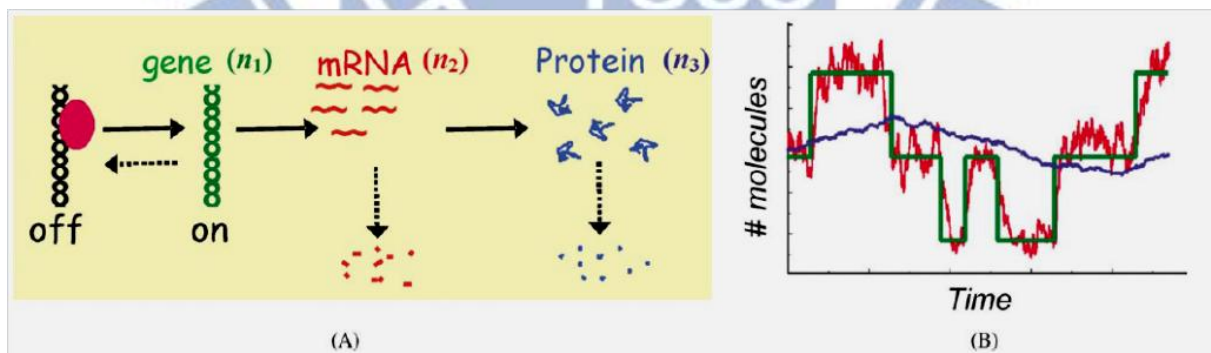
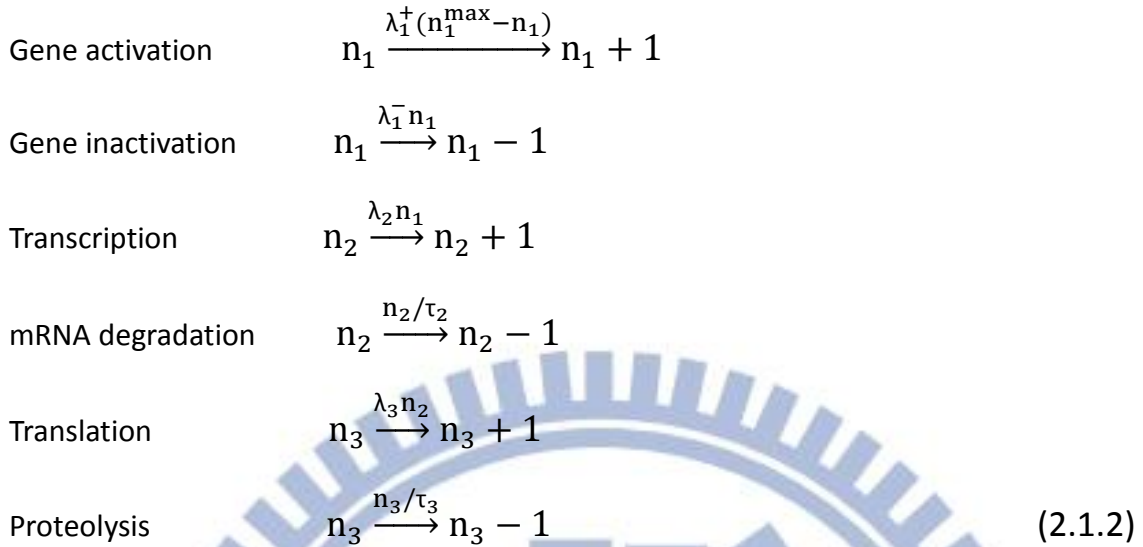


圖 2-1

(A) gene activation, transcription, translation, mRNA decay and proteolysis 的圖示

(B) 利用(2.1.2)的模擬結果，不同顏色曲線對應(A)中不同顏色物種，圖中可見 mRNA 的濃度相對於 protein 更快速的緊跟著 gene 的濃度做變化。

接著列出每個事件的反應式：



由上面六個反應式可分別得到以下 rate equations：

$$\begin{aligned}
 \frac{d\langle n_1 \rangle}{dt} &= \lambda_1^+(n_1^{\max} - \langle n_1 \rangle) - \lambda_1^- \langle n_1 \rangle = \lambda_1^+ n_1^{\max} - \langle n_1 \rangle / \tau_1 \\
 \frac{d\langle n_2 \rangle}{dt} &= \lambda_2 \langle n_1 \rangle - \langle n_2 \rangle / \tau_2 \\
 \frac{d\langle n_3 \rangle}{dt} &= \lambda_3 \langle n_2 \rangle - \langle n_3 \rangle / \tau_3
 \end{aligned} \tag{2.1.3}$$

其中參數 $\tau_1 = (\lambda_1^+ + \lambda_1^-)^{-1}$ 是一個基因活性變化的平均壽命， τ_2 、 τ_3 分別是 mRNA 和 proteins 的平均壽命。平均行為的動力學可以由化學反應式(2.1.3)立即推得，但想要了解其它的統計行為就必須透過 master equation：

$$\begin{aligned}
 & \frac{dP(n_1, n_2, n_3)}{dt} \\
 &= \lambda_1^+(n_1^{\max} - n_1 + 1)P(n_1 - 1, n_2, n_3) - \lambda_1^+(n_1^{\max} - n_1)P(n_1, n_2, n_3) \\
 &+ \lambda_1^-(n_1 + 1)P(n_1 + 1, n_2, n_3) - \lambda_1^- n_1 P(n_1, n_2, n_3)
 \end{aligned}$$

$$\begin{aligned}
& +\lambda_2 n_1 P(n_1, n_2 - 1, n_3) - \lambda_2 n_1 P(n_1, n_2, n_3) \\
& + (n_2 + 1)/\tau_2 P(n_1, n_2 + 1, n_3) - n_2/\tau_2 P(n_1, n_2, n_3) \\
& + \lambda_3 n_2 P(n_1, n_2, n_3 - 1) - \lambda_3 n_2 P(n_1, n_2, n_3) \\
& + (n_3 + 1)/\tau_3 P(n_1, n_2, n_3 + 1) - n_3/\tau_3 P(n_1, n_2, n_3)
\end{aligned} \tag{2.1.4}$$

由於(2.1.3)是 n_i 的線性微分方程式，所以可以解出隨時間變化的 n_i 。

接著利用在 stationary distribution 時 $dP/dt=0$ 可以解得 normalized variances[1]：

Proteins 引起的擾動：

$$\begin{aligned}
\frac{\sigma_3^2}{\langle n_3 \rangle^2} &= \underbrace{\frac{1}{\langle n_3 \rangle}}_{\text{Poisson}} + \underbrace{\frac{1}{\langle n_2 \rangle} \frac{\tau_2}{\tau_3 + \tau_2}}_{\text{Poisson One-step time-averaging}} + \underbrace{\frac{1 - P_{on}}{\langle n_1 \rangle} \frac{\tau_2}{\tau_2 + \tau_3} \frac{\tau_1}{\tau_1 + \tau_3} \frac{\tau_1 + \tau_3 + \tau_1 \tau_3 / \tau_2}{\tau_1 + \tau_2}}_{\text{Binomial Two-step time-averaging}}
\end{aligned} \tag{2.1.5}$$

在等式右邊的第一噪音項表明如果所有其他的細胞因素都是恆定的，proteins 個別出生與死亡事件的機率性依然會造成些微的擾動。除了上述影響外，擴散分子之間隨機碰撞的個別化學事件也會有噪音貢獻。第二項和第三項則是在反應 proteins 合成隨機變化的速度。

mRNA 引起的擾動：

$$\frac{\sigma_2^2}{\langle n_2 \rangle^2} = \underbrace{\frac{1}{\langle n_2 \rangle}}_{\text{Poisson}} - \underbrace{\frac{1 - P_{on}}{\langle n_1 \rangle}}_{\text{Binomial}} \underbrace{\frac{\tau_1}{\tau_2 + \tau_1}}_{\text{One-step time-averaging}} \tag{2.1.6}$$

(2.1.6)的第一項再次反應了 mRNA 固有出生和死亡的機率引起的些微噪音。第二項是來自基因激活的隨機變化。

genes 引起的擾動：

$$\frac{\sigma_1^2}{\langle n_1 \rangle^2} = \frac{1}{n_1^{\max}} \frac{\lambda_1^-}{\lambda_1^+} = \frac{1 - P_{\text{on}}}{\underbrace{\langle n_1 \rangle}_{\text{Binomial}}}, \quad \text{where } P_{\text{on}} = \frac{\langle n_1 \rangle}{n_1^{\max}} \quad (2.1.7)$$

(2.1.7) 的係數 $1 - P_{\text{on}}$ 來自於 active gene 的 binomial 分佈，而不是 Poisson 分佈。

在任何給定的平均下，Binomial 變量有相較於 Poissonian 變量更小的相對擾動，這使的只需要少數量且大部分具有活性的基因，而不一定要有大幅的擾動， $P_{\text{on}} \approx 1$ 。

2.2 利用 Dual-Reporter Method 來區分訊號

只要在一個實驗裡定義了環境與系統，我們就可以將 Total variance 拆解成來自外部的噪音 Extrinsic 與來自內部的噪音 Intrinsic。

Paulsson 等人選擇使用 Dual-Reporter 的方法來拆解噪音，首先我們在單一細胞中加入一個 reporter X，之後再加入第二個與第一個完全相同但是各自獨立且不相關的 reporter Y，在這裡兩個 reporters 完全沒有任何區別且不會影響對方，其中細胞和 reporters 分別定義為環境和系統，在不考慮實驗誤差對內部與外部噪音觀測的影響下，首先可以將 Total variance 拆解成兩項：

$$\eta_{\text{tot}}^2 \equiv \frac{\sigma_x^2}{\langle x \rangle^2} = \eta_{\text{int}}^2 + \eta_{\text{ext}}^2 \quad (2.2.1)$$

其中兩個 reporters 之間的 Covariance 宣告為外部的噪音：

$$\eta_{\text{ext}}^2 = \frac{\text{Cov}(x,y)}{\langle x \rangle \langle y \rangle} \quad (2.2.2)$$

由於內部與外部噪音有明確的物理意義，上面的方程式不能視為定義而只是一種宣告。為了確認 dual-reporter 可用來以區別內外噪聲，我們必須先證明總的噪音根據 Eq(2.2.1) 的分解後，確實就是一般的內外噪聲。

2.3 利用當下環境的狀態來分解噪音

首先，利用當下環境的狀態為條件，總噪音可以如下推導：

$$\begin{aligned}
& \sigma_x^2 \\
&= \sum_{x,z} x^2 p(x,z) - (\sum_{x,z} xp(x,z))^2 \\
&= \sum_z \sum_x x^2 p \sigma_x^2(x|z) p(z) - (\sum_z \sum_x xp(x|z)p(z))^2 \\
&= \sum_z \sum_x x^2 p(x|z)p(z) - \sum_z (\sum_x xp(x|z))^2 p(z) \\
&+ \sum_z (\sum_x xp(x|z))^2 p(z) - (\sum_z \sum_x xp(x|z)p(z))^2 \\
&= \sum_z \sigma_{x|z}^2 p(z) + \sum_z \langle x|z \rangle^2 p(z) - (\sum_z \langle x|z \rangle p(z))^2 \\
&= \underbrace{\langle \sigma_{x|z}^2 \rangle}_{\text{unexplained by } Z} + \underbrace{\sigma_{\langle x|z \rangle}^2}_{\text{explained by } Z} \tag{2.3.1}
\end{aligned}$$

其中 $\langle \sigma_{x|z}^2 \rangle$ 表示在當下出現某 Z 條件下 X variance 的平均， $\sigma_{\langle x|z \rangle}^2$ 則表示在當下出現某 Z 條件下 x 平均的 variance。

在上述推導中，(2.3.1) 數學上等效於之前利用 dual-reporters method 所得出的(2.2.1)。

[2]

這種拆解的方式也最常被用來解釋內外噪聲的影響。

接著我們考慮一個外部變數 Z 的擾動會影響內部變數 X 的簡單模型：

Extrinsic dynamics	Intrinsic dynamics	
$Z \xrightarrow{\lambda_{\text{ext}}} Z + 1$	$X \xrightarrow{\lambda_{\text{int}}^Z} X + 1$	(2.3.2)
$Z \xrightarrow{\beta_{\text{ext}}^Z} Z - 1$	$X \xrightarrow{\beta_{\text{int}}^X} X - 1$	

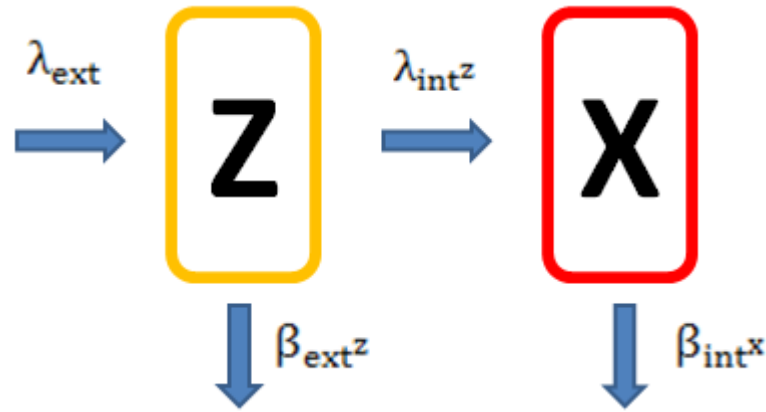


圖 2-2

其中 Z 與 X 分別為外部與內部變數， λ_{ext} 、 λ_{int}^z 、 β_{ext}^z 、 β_{int}^x 分別為反應速度常數並也對應到躍遷機率。

上述模型在穩態中(stationary)中，我們可以得到此模型 Variance 精確的分解：

$$\eta_{tot}^2 = \underbrace{\frac{1}{\langle x \rangle}}_{\eta_{int}^2} + \frac{1}{\langle z \rangle} \frac{\tau_z}{\tau_x + \tau_z} \quad (2.3.3)$$

其中 $\tau_z = 1/\beta_{ext}$ 和 $\tau_x = 1/\beta_{int}$ 是環境和內部相應的生命周期。[2]

由(2.3.3)可以知道總的噪音確實被分成了內部與外部項，且內部項中不含有外在環境擾動的影響，外部項中也不含有內部隨機性的影響。

為了看出(2.3.3)即為(2.3.1)的分解，必須計算(2.3.1)，這本來是項困難工作。但是利用 Paulsson 等人發展的遞迴關係，(2.3.1)裏分解的兩項可確切求出：

$$\frac{\sigma_{\langle x|z \rangle}^2}{\langle x \rangle^2} = \eta_{ext}^2 \frac{\tau_z}{\tau_x + \tau_z}$$

$$\frac{\langle \sigma_{x|z}^2 \rangle}{\langle x \rangle^2} = \eta_{int}^2 + \eta_{ext}^2 \left(1 - \frac{\tau_z}{\tau_x + \tau_z}\right) \quad (2.3.4)$$

和精確分解(2.3.3)比較發現，如果依當下環境狀態的方式(2.3.4)分解，分解的結果會誇大

內部噪音而低估了外部噪音。

只有在 $\tau_z \gg \tau_x$ 時，(2.3.4)才會回歸到(2.3.3)的精確解，。也就是環境的擾動在無限緩慢的極限下，依當下環境狀態的分解方式會趨近於精確解，但是在自然情況下很少有細胞是緩慢甚至靜止的，所以依當下環境狀態的分解方式並無法精確的分解出在動態環境下的噪音。

由此例子我們知道如果想要得到精確的分解，並不能只考慮當下的環境狀態，而必須考慮所有的歷史過程。

2.4 利用環境的歷史來分解噪音

我們將(2.3.1)中原本對當下環境狀態的條件全部更換成對環境所有歷史的條件：

$$\sigma_X^2 = \underbrace{\langle \sigma_{X_t | Z[0,t]}^2 \rangle}_{\sigma_{int}^2} + \underbrace{\sigma_{\langle X_t | Z[0,t] \rangle}^2}_{\sigma_{ext}^2} \quad (2.4.1)$$

其中內部項 σ_{int}^2 是 X_t 在一群共享所有可能路徑的環境歷史 $Z[0,t]$ 的細胞 variance 的平均，外部項則是所有可能的歷史路徑 $Z[0,t]$ 下 X_t 的平均值的 variance。

如果利用(2.4.1)考慮所有歷史過程的方式來分析之前(2.3.4)的簡單模型，即可得到(2.3.3)的正確內外噪聲分解。

如此定義的話，外部噪音 σ_{ext}^2 就可被解釋為平均值對時間的變化，而 σ_{int}^2 就對應於平均的周圍。

另外，如果系統的環境變量 Z 是遍歷(ergodic)的，那一群 Z 的平均將可以等效於單一 Z 軌跡對時間的平均。

$$\langle X_t | Z[0,t] \rangle_{z[0,t]} = \overbrace{\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{j=1}^m x^j(t) dt}_{\text{ensemble average } \langle X_t | Z[0,t] \rangle} \quad (2.4.2)$$

其中上標 j 代表在同一條路徑 $Z[0,t]$ 下出現的 m 個 X_t 。

因此可用(2.4.2)計算(2.4.1)當中的某些項(見 chap 4.2.1)。

在分解 (2.4.1) 時，原則上需要觀察在同一細胞中大量的 reporters 來獲得。但實際上利用兩個 reporters 就可以觀察到相同結果。在相同環境下，如果系統 X、Y 是兩個完全相同但不相關的 reporter，在不會影響環境也不會影響彼此的狀況下，他們會滿足以下統計行為

$$\langle X_t Y_t | Z[0, t] \rangle = \langle X_t | Z[0, t] \rangle \langle Y_t | Z[0, t] \rangle$$

$$\langle X_t | Z[0, t] \rangle = \langle Y_t | Z[0, t] \rangle$$

在任何非線性且多元的過程中，兩個完全相同且不相關的 reporter 之間的共方差對應到 (2.4.1) 中定義的外部噪音：

$$\begin{aligned} \text{Cov}(X, Y) &= \langle\langle X_t Y_t | Z[0, t] \rangle\rangle - \langle\langle X_t | Z[0, t] \rangle\rangle \langle\langle Y_t | Z[0, t] \rangle\rangle \\ &= \langle\langle X_t | Z[0, t] \rangle \langle Y_t | Z[0, t] \rangle \rangle - \langle\langle X_t | Z[0, t] \rangle\rangle^2 \\ &= \langle\langle X_t | Z[0, t] \rangle^2 \rangle - \langle\langle X_t | Z[0, t] \rangle\rangle^2 \equiv \sigma_{\text{ext}}^2 \quad (2.4.3) \end{aligned}$$

其中外括號是對 Z 做平均。

所以在分解總噪音時可用計算兩個 reporters 間的共方差來代替觀察無限多的 reporters。

下圖為動態系統中內部噪音與外部噪音的物理意義。

圖中紅跟藍線為環境隨時間演化(灰色斜線)下，系統跟隨的隨機路徑，綠色實線為這些路徑的平均，外部噪音為綠色實線的變異性(variability)；內部噪音則是圖中亮綠色的範

圍，意即紅跟藍線對綠線變異性的平均。

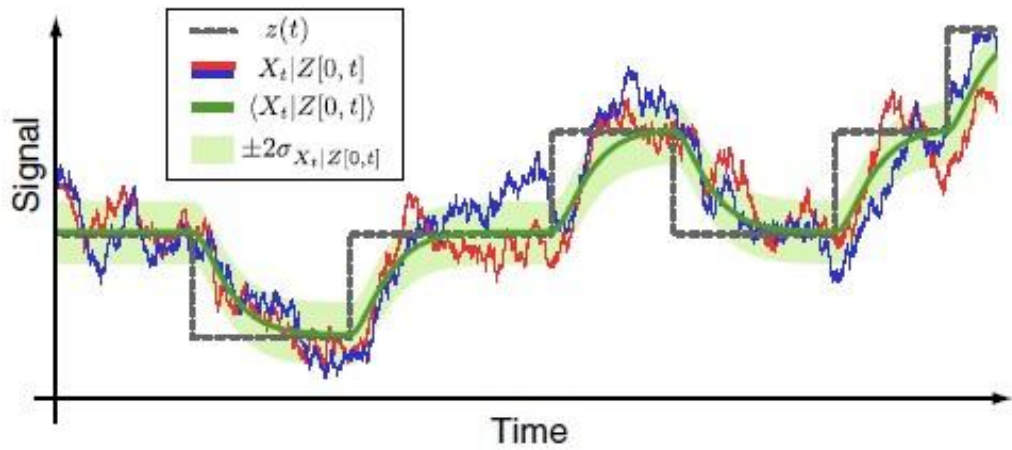


圖 2-3

2.5 KL_Divergence

在機率理論與資訊理論裡，Kullback–Leibler divergence 是 P 和 Q 兩個機率分佈之間差異的非對稱測量，在 P 和 Q 是離散的機率分布時，從 P 到 Q 的 KL divergence 定義為：

$$D_{KL} = \sum_i \ln \left(\frac{P(i)}{Q(i)} \right) P(i)$$

換句話說，它是以機率分布 P 為基準，機率分布 P 和 Q 之間的對數差的期望值。

KL divergence 只在 P、Q 各自總和為 1 時有定義且 $D_{KL}(P||Q) \geq 0$ 。

第3章 系統架構

3.1 系統

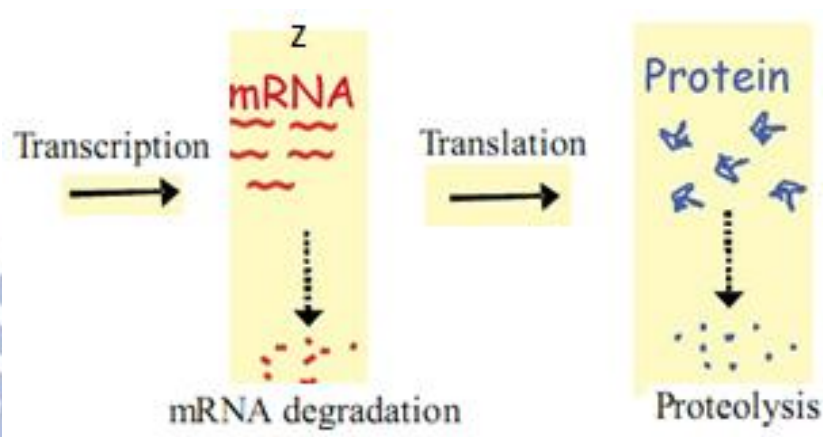


圖 3-1

上圖是我們的研究系統，為 mRNA 轉換成 Protein 的過程，其中四個主要反應過程為：
 (i)Transcription，在這裡將它簡化為一個常數與 mRNA 跟 Protein 的濃度皆無關；(ii)mRNA degradation，mRNA 的降解作用會使其濃度下降；(iii)Translation，mRNA 的濃度會影響 Protein 濃度的增加及 (iv)Proteolysis，為 proteins 水解造成 Protein 濃度的下降。

這四個過程對應到四條化學反應方程式：

	Extrinsic Dynamics	Intrinsic Dynamics	
Transcription	$Z \xrightarrow{\lambda_{\text{ext}}} Z + 1$	$X \xrightarrow{\lambda_{\text{int}}^z} X + 1$	Translation
mRNA degradation	$Z \xrightarrow{\beta_{\text{ext}}^z} Z - 1$	$X \xrightarrow{\beta_{\text{int}}^x} X - 1$	Proteolysis

其中 λ_{ext} 、 λ_{int}^z 、 β_{ext}^z 、 β_{int}^x 為他們所對應到的反應速度常數。

由這些反應方程式得到下列 rate equations。

$$\begin{aligned}\frac{dz}{dt} &= \lambda_{\text{ext}} - \beta_{\text{ext}z} * z \\ \frac{dx}{dt} &= \lambda_{\text{int}z} * z - \beta_{\text{int}x} * x\end{aligned}\quad (3.1.1)$$

可知其中 Z 的濃度變化只與自己有關，但 X 濃度變化則與 mRNA 與 Protein 都有關係。

3.2 數值模擬方法

我們先令 N_z 、 N_x 分別代表 Z 與 X 的粒子個數，並將速度常數 λ_{ext} 、 $\lambda_{\text{int}z}$ 、 $\beta_{\text{ext}z}$ 、 $\beta_{\text{int}x}$ 與一組 N_z 、 N_x 的初始條件設定好。

利用 master equation，每隔一段極微小時間 Δt ，讓系統在 Transcription、Translation、mRNA degradation、Proteolysis 以及不反應中，依照他們個別的機率去擇一進行。因為極微小，在每 Δt 中只會有一個粒子的跳動或是沒有發生跳動這兩種可能。

利用這種方式，在經過一段時間後觀察 N_z 、 N_x 隨時間的變化，即為 Z 與 X 對時間演化的結果

第4章 研究結果

4.1 利用當下環境的狀態來分解噪音

在這一節中我們將利用三種不同的方式來驗證第二章中利用環境的狀態來分解噪音的理論是否符合。首先，我們用微分方程式求得 X 、 Z 的平均值，將他們跟已知的 **rate constant** 代入(2.3.4)中可得 $\sigma_{\langle x|z \rangle}^2$ (Unexplained by Z) 及 $\langle \sigma_{x|z}^2 \rangle$ (Explained by Z) 的理論值。其次，由數值模擬方法分別利用 **time average** 與 **ensemble average** 兩種不同的演算方式來計算 $\sigma_{\langle x|z \rangle}^2$ 及 $\langle \sigma_{x|z}^2 \rangle$ ，最後再比較三種不同方式得到的結果是否一致。

4.1.1 使用微分方程式求變異數解

首先，我們利用數值模擬及直接解微分方程，驗證之前所提到「利用當下環境的狀態來分解噪音」的方法，利用(2.3.3)中 η_{ext}^2 、 η_{int}^2 的定義代入(2.3.4)可以推得：

$$\underbrace{\sigma_{\langle x|z \rangle}^2}_{\text{Unexplained by Z}} = \left(\left(\frac{1}{\langle z \rangle} \frac{\tau_z}{\tau_x + \tau_z} \right) * \frac{\tau_z}{\tau_x + \tau_z} \right) * \langle x \rangle^2 \quad (4.1.1)$$

$$\underbrace{\langle \sigma_{x|z}^2 \rangle}_{\text{Explained by Z}} = \left(\frac{1}{\langle x \rangle} + \left(\frac{1}{\langle z \rangle} \frac{\tau_z}{\tau_x + \tau_z} \right) \left(1 - \frac{\tau_z}{\tau_x + \tau_z} \right) \right) * \langle x \rangle^2$$

$$\tau_z = 1/\beta_{\text{ext}^x} \quad , \quad \tau_x = 1/\beta_{\text{int}^x}$$

利用(3.1.1)式解出的 $\langle z \rangle$ 與 $\langle x \rangle$ 代入(4.1.1)式即可得到 $\sigma_{\langle x|z \rangle}^2$ 、 $\langle \sigma_{x|z}^2 \rangle$ 理論上的解。

在 **rate constants** 為 $\lambda_{\text{ext}} = 15$ 、 $\lambda_{\text{int}^z} = 175$ 、 $\beta_{\text{ext}^z} = 4$ 、 $\beta_{\text{int}^x} = 410$ 時， $\sigma_{\langle x|z \rangle}^2$ 及 $\langle \sigma_{x|z}^2 \rangle$ 分別為 1.594106 及 0.662906093。

4.1.2 使用 time average 求變異數

由數值模擬方法產生隨時間演化的 Z 與 X (下圖左)。

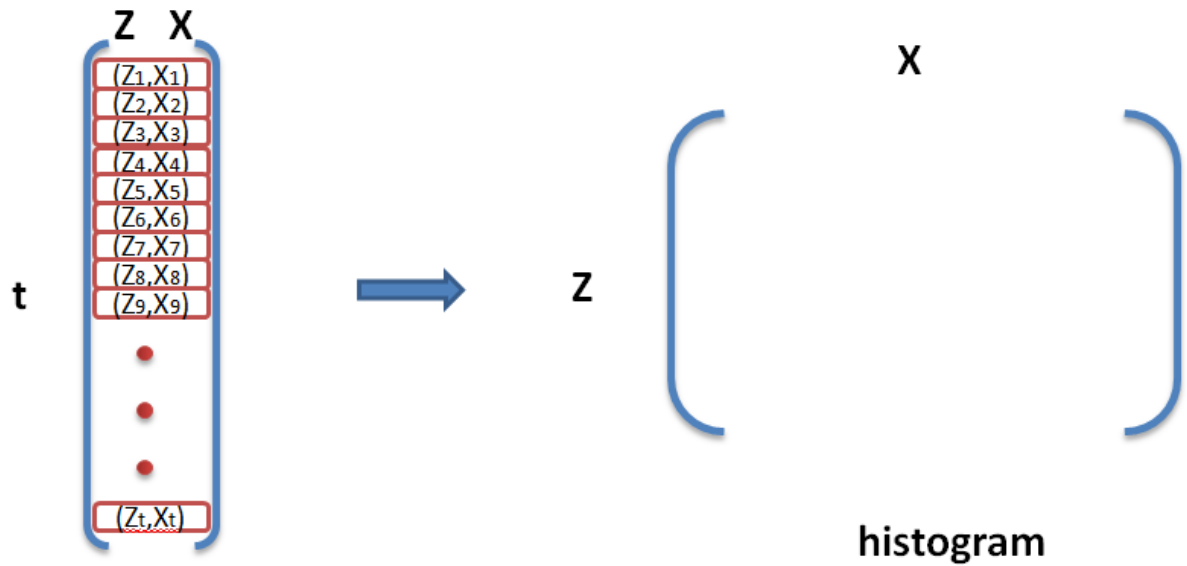


圖 4-1

統計不同時間的不同組合 (Z, X) 出現的次數就可得到不同 (Z, X) 值的出現次數統計表(上圖右), 由該統計圖也可求得不同狀態 Z 出現的機率以及在不同條件狀態 Z 下 X 出現的機率。

圖 4-2 為在一次長時間實驗下統計 (Z, X) 出現次數所做的立體圖。

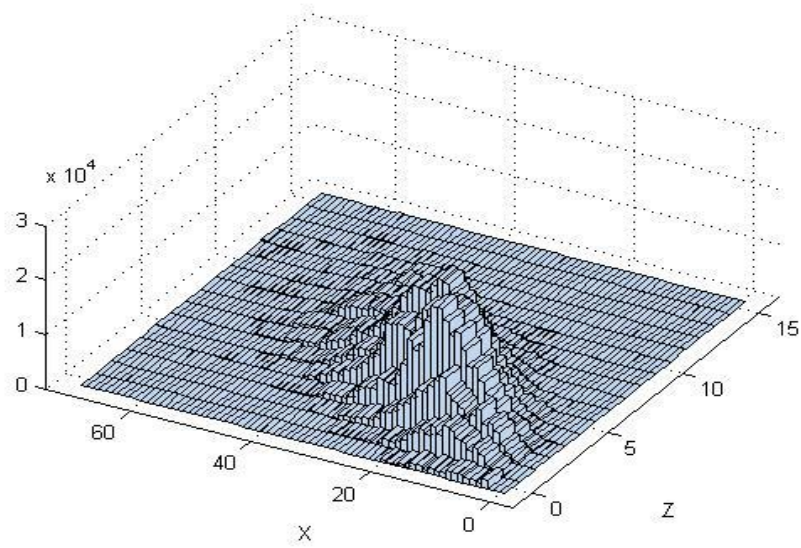


圖 4-2

利用這些資訊及 (2.3.1) 式，可計算在不同條件狀態 Z 下， X 平均的變異數

$\sigma^2_{\langle X | Z \rangle}$ (Explained by Z) 與在不同條件狀態 Z ， X 變異數的平均 $\langle \sigma^2_{X | Z} \rangle$ (Unexplained by Z)。

所得結果如下表

t(萬)	110	120	130	140	150	160	170	180	190	200
Explained by Z	0.689	0.657	0.664	0.571	0.682	0.667	0.673	0.642	0.741	0.671
Unexplained by Z	1.578	1.629	1.655	0.638	1.637	1.506	1.658	1.601	1.638	1.682

將它畫成圖，為下圖中藍色曲線，而圖中紅線為 4.1.1 節算出的理論解。

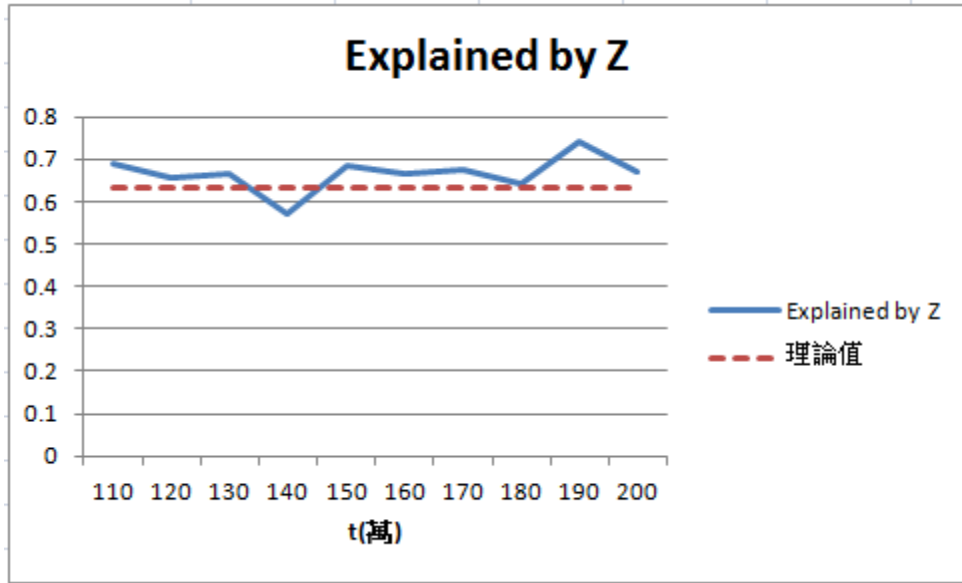


圖 4-3

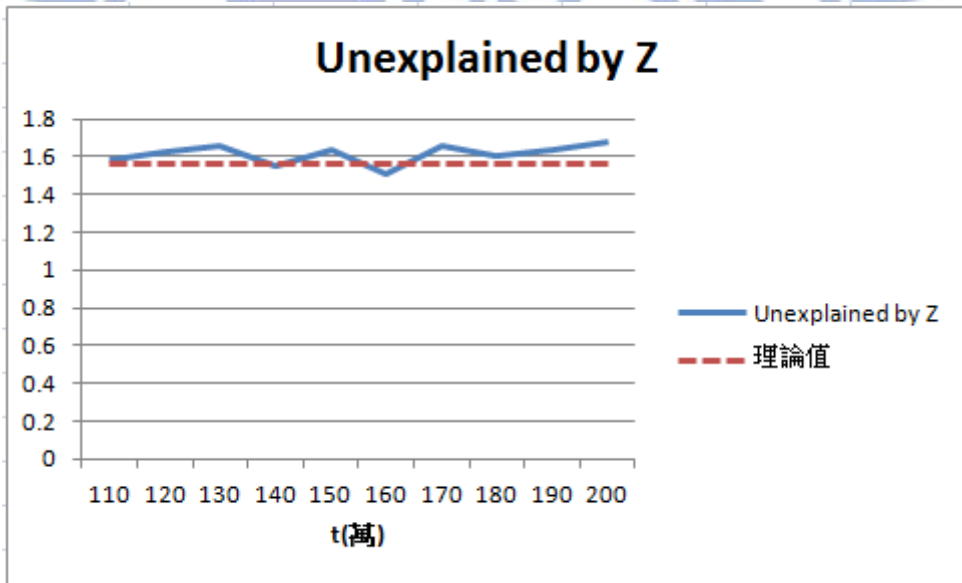


圖 4-4

上述數據與圖表表示了所得到的結果隨著實驗時間的增加是相當穩定的。相互比較之後發現模擬算出的變異數和 4.1.1 節算出的理論解相當一致。

4.1.3 使用 ensemble average 求變異數

由數值模擬方法產生多組實驗並讓它們各自隨時間演化 Z 與 X(下圖左)。

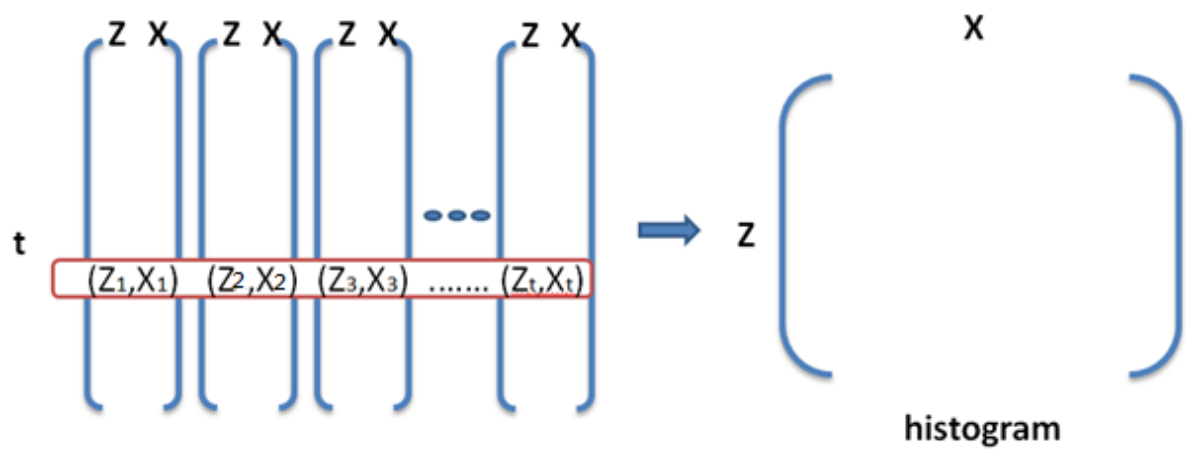


圖 4-5

統計在某一時間下所有組(Z,X)亦可得到 X、Z 的統計圖(上圖右)，由該統計圖表可求得不同狀態 Z 出現的機率以及在不同條件狀態 Z 下 X 出現的機率。圖 4-6 為對不同實驗在同一時間下統計(Z,X)出現次數所做的立體圖

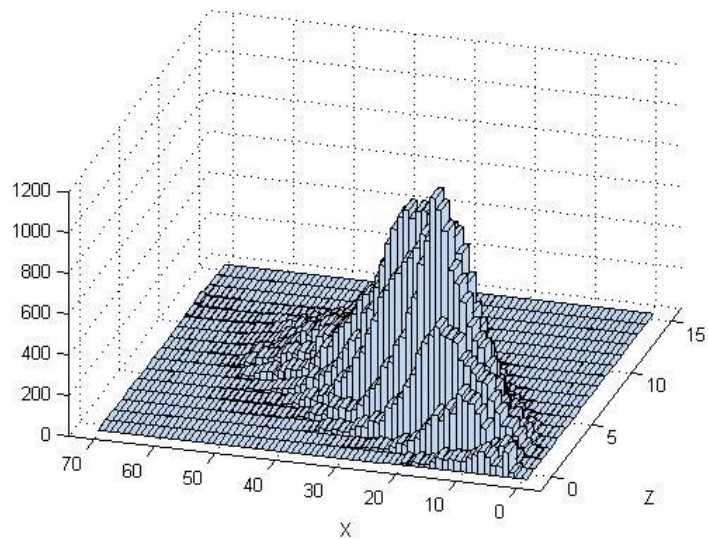


圖 4-6

利用這些資訊及 (2.3.1) 式，可計算不同條件狀態 Z 下 X 平均的變異數

$\sigma^2_{\langle X | Z \rangle}$ (Unexplained by Z) 與在不同條件狀態 Z 下，X 變異數的平均 $\langle \sigma^2_{X | Z} \rangle$ (Explained by Z)。所得結果如下表

t(萬)	110	120	130	140	150	160	170	180	190	200
Explained by Z	0.621	0.616	0.621	0.614	0.598	0.683	0.687	0.621	0.654	0.625
Unexplained by Z	1.520	1.624	1.540	1.581	1.562	1.589	1.611	1.562	1.575	1.567

如在圖 4-3 及 4-4 做的，上表 data 在圖 4-7 及 4-8 中為藍色曲線，而紅線為 4.1.1 節算出的理論解。

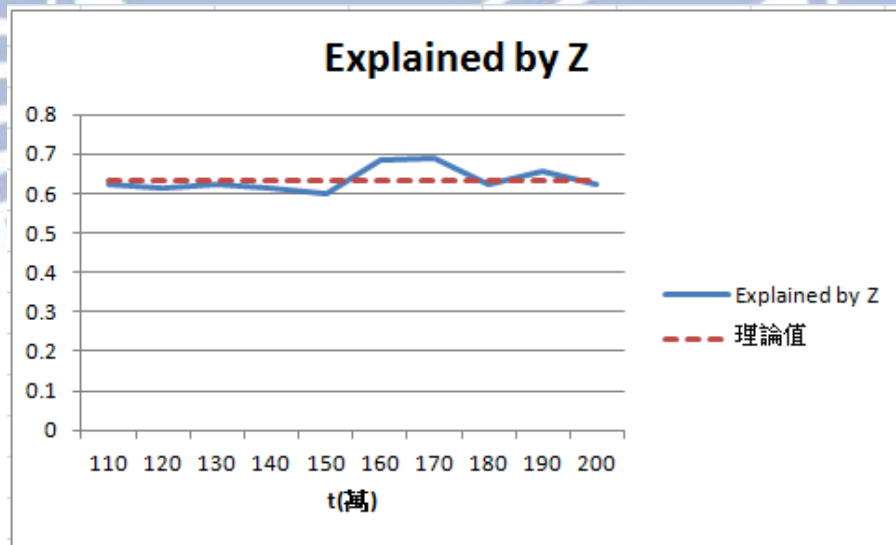


圖 4-7

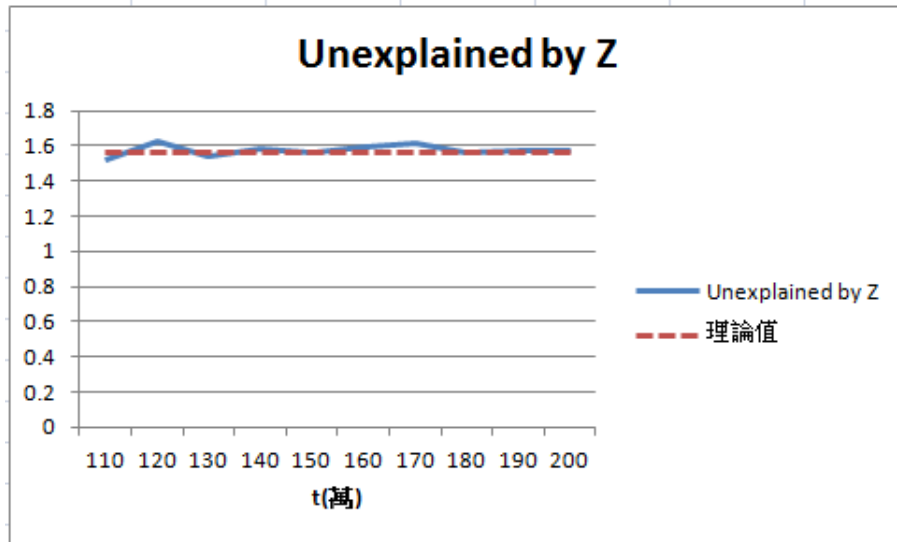


圖 4-8

上述數據與圖表表示了所得到的結果隨著實驗時間的增加是相當穩定的。相互比較之後發現模擬算出的變異數和 4.1.1 節算出的理論解相當一致。

4.2 利用環境的歷史來分解噪音

在這節中我們將先利用 ergodic 下的近似方法來對長時間環境歷史做平均，接著取一個 ensemble 的環境 Z 來做平均。

再藉由改變不同的 rate constants 來了解環境影響系統的強弱也會影響到環境歷史過程的重要性。

4.2.1 使用微分方程式求變異數解

當我們使用「利用環境的歷史來分解噪音」的方法時，能正確的拆解噪音，由(2.3.5)可知：

$$\eta_{\text{int}}^2 = \frac{\langle \sigma_{x_t | Z[0,t]}^2 \rangle}{\langle x \rangle^2} = \frac{1}{\langle x \rangle} \quad (4.2.1)$$

$$\eta_{\text{ext}}^2 = \frac{\sigma_{x_t | Z[0,t]}^2}{\langle x \rangle^2} = \frac{1}{\langle z \rangle} \frac{\tau_z}{\tau_x + \tau_z}$$

$$\tau_z = 1/\beta_{\text{ext}x} \quad , \quad \tau_x = 1/\beta_{\text{int}x}$$

利用(3.1.1)式解出的 $\langle z \rangle$ 與 $\langle x \rangle$ 代入 (4.2.1) 式即可得到 $\langle \sigma_{x_t | Z[0,t]}^2 \rangle$ 、 $\sigma_{x_t | Z[0,t]}^2$ 理

論上的解。

在 rate constants 為 $\lambda_{\text{ext}} = 15$ 、 $\lambda_{\text{int}^z} = 175$ 、 $\beta_{\text{ext}^z} = 4$ 、 $\beta_{\text{int}^x} = 410$ 時， $\sigma_{\langle X_t | Z[0,t] \rangle}^2$ 及 $\langle \sigma_{X_t | Z[0,t]}^2 \rangle$ 分別為 1.561158 及 0.632397。

4.2.2 在 ergodic 條件下求變異數

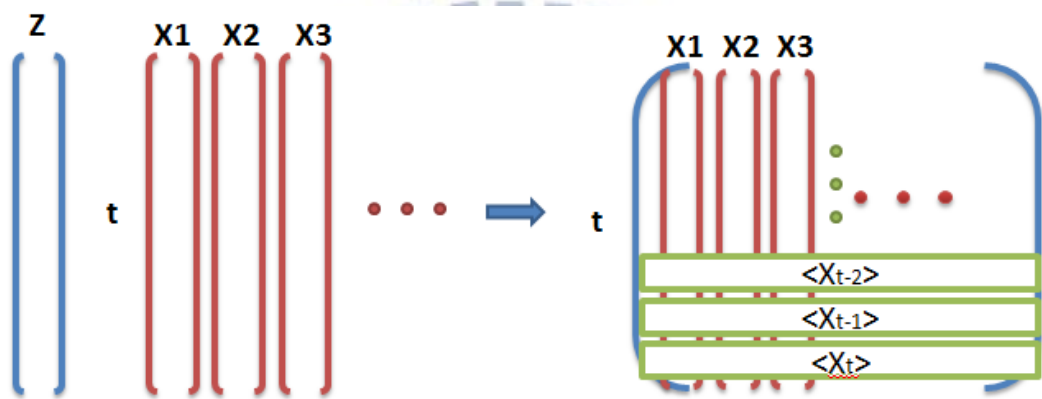


圖 4-9

由數值模擬方法先產生一條隨時間演化的 Z 隨機路徑，在起始的 Z 值下不同 X 值的出現會有一個機率分布，以該分佈隨機選取一堆 X ，和 Z 的初始值搭配成一個 ensemble 裏個別系統的初始 (Z, X) ，讓他們在同一 Z 的歷史環境下隨時間做演化(上圖左)，統計在該 ensemble 下時間 t 時 X_t 出現某值的機率(上圖右)，算出 X_t 的平均 $\langle X_t | Z[0, t] \rangle$ (圖 4-9 右的綠色橫列平均)，接者利用如(2.3.1)的基本定義即可算出內外噪聲的變異數：

由(2.4.1)

$$\begin{aligned} \sigma_X^2 &= \langle X^2 \rangle - \langle X \rangle^2 \\ &= \underbrace{\langle \langle X_t^2 | Z[0, t] \rangle \rangle - \langle \langle X_t | Z[0, t] \rangle^2 \rangle}_{\langle \sigma_{X_t | Z[0, t]}^2 \rangle} + \underbrace{\langle \langle X_t | Z[0, t] \rangle^2 \rangle - \langle \langle X_t | Z[0, t] \rangle \rangle^2}_{\sigma_{\langle X_t | Z[0, t] \rangle}^2} \end{aligned}$$

$$\sigma_{\text{ext}}^2 = \langle \langle X_t | Z[0, t] \rangle^2 \rangle - \langle \langle X_t | Z[0, t] \rangle \rangle^2$$

$$\sigma_{\text{int}}^2 = \langle \langle X_t^2 | Z[0, t] \rangle \rangle - \langle \langle X_t | Z[0, t] \rangle^2 \rangle \quad (4.2.2)$$

其中，外括號是對 Z 做平均。

在 ergodic 假設底下，(4.2.2)中外括號可由(2.4.2)的長時間平均取代。其中(2.4.2)裏長時間積分以對所有圖 4-9 右的綠橫條求和取代。

在此假設下算出內外噪聲的變異數如下表：

t(萬)	10	20	30	40	50	60	70	80
Extrinsic	0.596	0.569	0.712	0.627	0.691	0.690	0.675	0.630
Intrinsic	1.558	1.606	1.545	1.719	1.642	1.604	1.677	1.582

上表 data 在圖 4-10 及 4-11 中為藍色曲線，而紅線為 4.1.1 節算出的理論解。

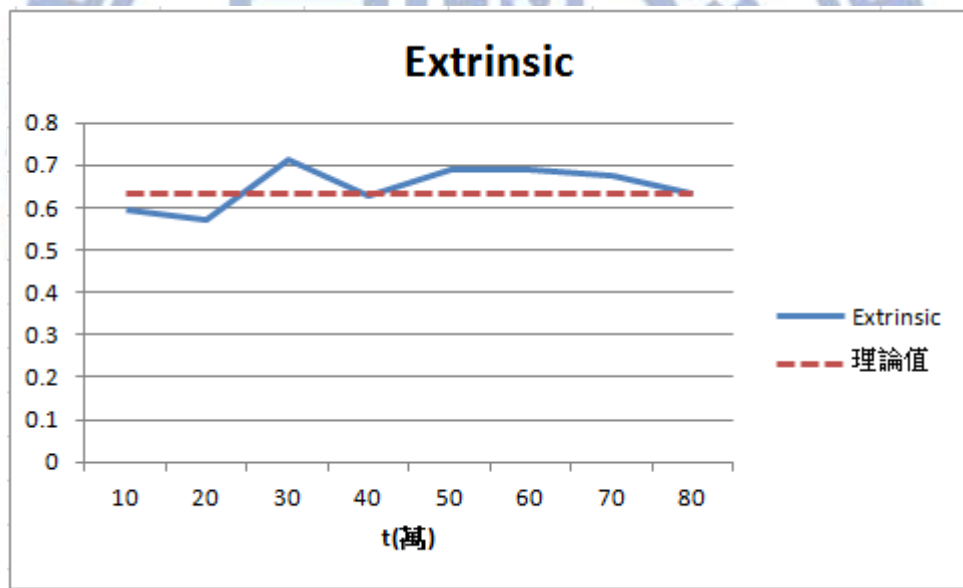


圖 4-10

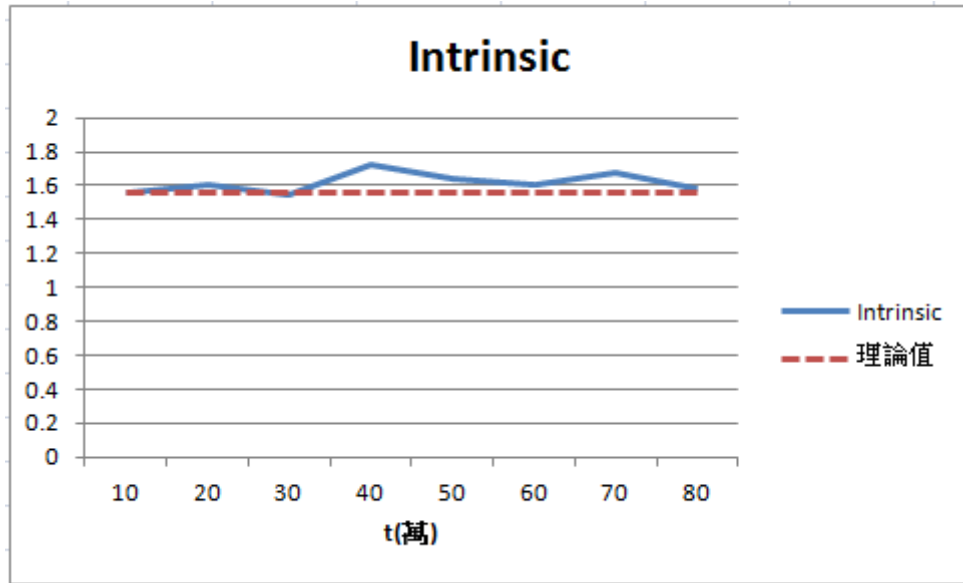


圖 4-11

圖中 X 軸為不同的演化時間，可以得知結果是穩定的。

比較後發現與 4.2.1 節中求出來的理論值是一致的

4.2.3 取不同歷史長短求 Extrinsic & Intrinsic 變異數

同樣先由數值模擬方法，讓 Z 與 X 隨時間演化一段時間至穩態之後，再對(Z,X)出現機率做統計：

首先，將每 t 間距的 Z 與 X 挑出，將他們視作經過 t 時間演化的多組實驗，利用這些資訊及(4.2.2)，我們可以求出 Intrinsic 與 Extrinsic 的變異數。和 4.2.2 節不同的是(4.2.2)裏對 Z 的平均的外括號現在是以上述多組實驗來求平均，不同於之前的 ergodic 近似下用(2.4.2)裏的時間平均取代 Z 的平均。計算結果呈現於下兩圖：

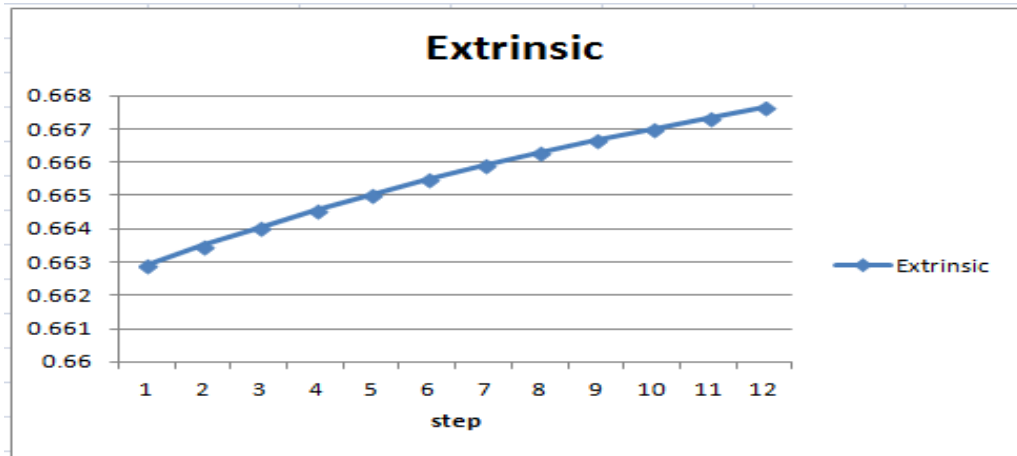


圖 4-12

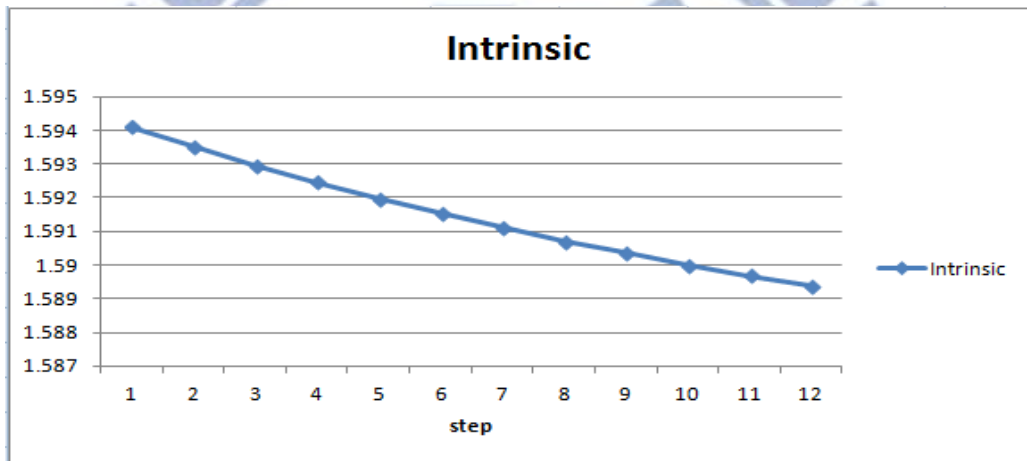


圖 4-13

由圖可發現，當 $t=1$ 時所得到的內外噪聲變異數就跟 4.1 節以當下環境的狀態 Z 來分解噪聲是一樣的。隨著 t 的增加該變異數有些微變化，顯示計算 Z 歷史的重要性。另外，intrinsic 圖告訴我們，如(2.3.4)所言，只考慮當下 Z 底下的 X 會誇大內噪聲。然而，雖然內外噪聲變異數隨 t 變化，但幅度不大，表示 4.1.2 及 4.1.3 的當下環境估算還算夠好。這跟系統有關，在目前這組速率常數，變異數計算確實不太需要知道 Z 的歷史，但對一般速率常數， Z 的歷史是需要的。

4.2.4 不同速度常數的 Extrinsic 及 Intrinsic 變異數

在這裡我們利用 4.2.3 的方法，在不同速率常數下觀察 Extrinsic 與 Intrinsic 變異數計算得到的值隨著考量的歷史長度 t 的改變。

下圖為四種不同速率常數，取不同歷史長短(step)的 Z ，計算系統變異數，不同線的速率常數表示為 $\lambda_{\text{ext}}-\beta_{\text{ext}z}-\lambda_{\text{int}z}-\beta_{\text{int}x}$ 。不同速率常數下的變異數不同，為方便比較，將它們在 $t=1$ 的值設為零。

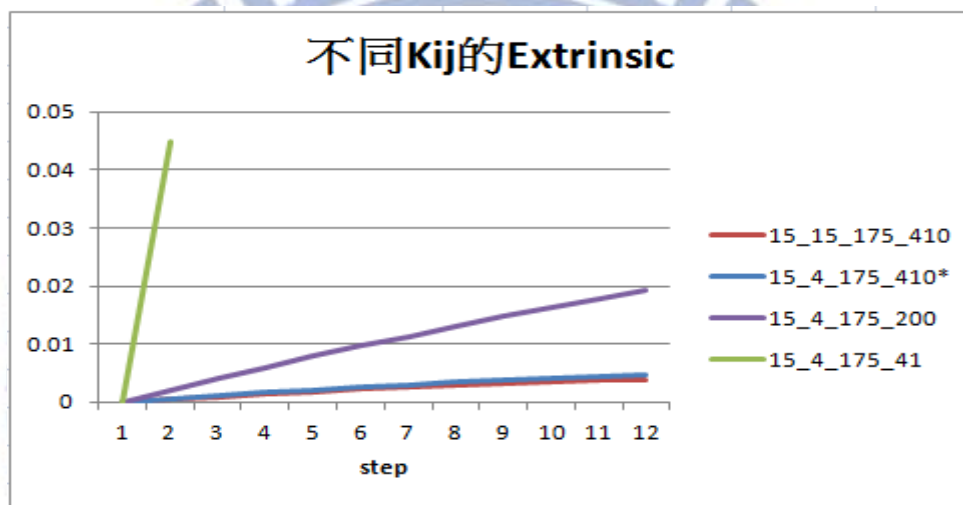


圖 4-14

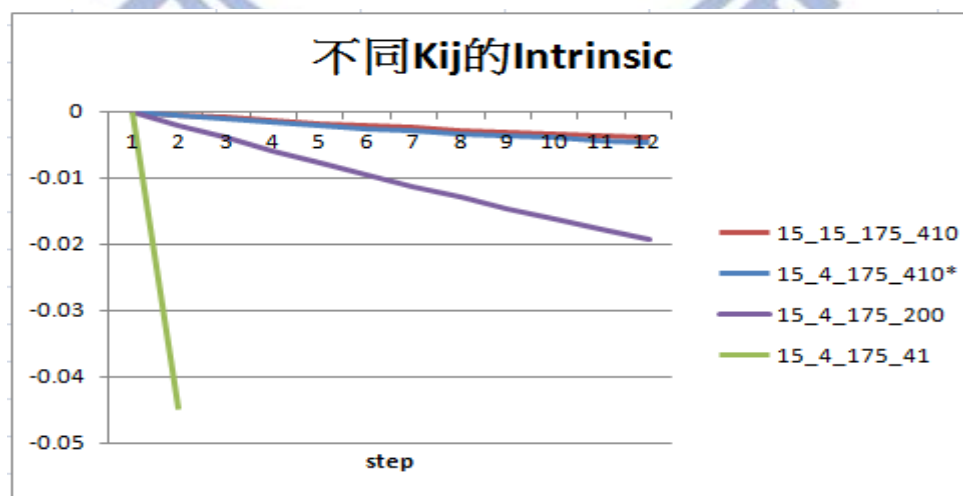


圖 4-15

不同速率常數，系統 relax 的速度不同。

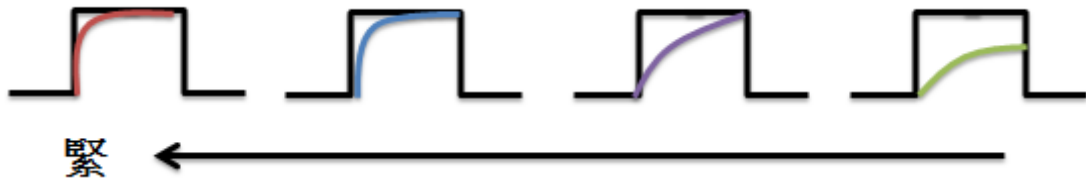


圖 4-16

在圖 4-16 中，黑線表示 mRNA 的變化，不同顏色的線表示不同速率常數下的 protein 跟著 mRNA 的變化，其中顏色對應圖 4-15 跟 4-16 中不同速率常數的顏色。

當 protein 能緊緊的跟著 mRNA 變化時，protein relaxation 快(如紅線)，。反之，當 mRNA 做了改變時 protein 也不會有太大的改變，protein relaxation 慢(如綠線)。

比較圖 4-14、4-15、4-16 後發現當 protein 能跟 mRNA 跟的越緊時(紅線)，隨著歷史記錄的增加(增大 t)，估算得到的 Extrinsic 及 Intrinsic 變異數只會有些微的變化。但是當 protein 跟得越不緊時(綠線)，估算得到的變異數的變化就會越大。也就是當 protein relaxation 快(如紅線)，環境歷史對變異數的估算就越不重要，但如果 protein relaxation 慢(如綠線)，環境歷史變得相當重要，因此必須盡可能的了解所有的環境歷史才能對擾動的來源做正確的分析。

另外在圖 4-14 與圖 4-15 中的藍、紫、綠線是固定了其他的速率常數，藉由只改變 τ_x 來觀察變異數的變化，其中藍線的 τ_z 相對於 τ_x 最大，由(2.3.1)與公式(2.3.2)知道藍線在以當下環境為條件拆解的內外噪音分別為 1.60736 及 0.66988，在以環境的歷史為條件下拆解的內外噪音分別為 1.60083 及 0.67641，可以知道在 $\tau_z \gg \tau_x$ 時，在利用當下環境為條件做拆解時就已經相當精確。

4.3 X 與不同時間之前的 Z 的關聯

在這節我們分別使用「KL-Divergence」與「variance norm」來了解歷史 t 要取多長，

$\langle \sigma_{x_t}^2 | z_{[0,t]} \rangle$ 與 $\sigma_{\langle x_t | z_{[0,t]} \rangle}^2$ 的估算才會收斂到一個令人滿意的值，

分析數據的方法為先取某時刻之前 t 時間間距之內的 Z 路徑(圖 4-17 右藍橫帶)，再取該時刻之前 $t+1$ 時間間距之內的 Z 路徑(圖 4-17 中藍橫帶)，在圖 4-17 的例子， $t+1$ 時刻的 Z 取值出現了 1, 2, 3 三種可能。以 $t+1$ 長度的 Z 數列為條件，觀察最後一個 Z 出現時的 X 值是多少，統計該 X 取值的 histograms，畫於圖 4-17 左。其中橫軸表示在統計過程中有出現的 X 值，縱軸則為不同 X 值所出現的機率。

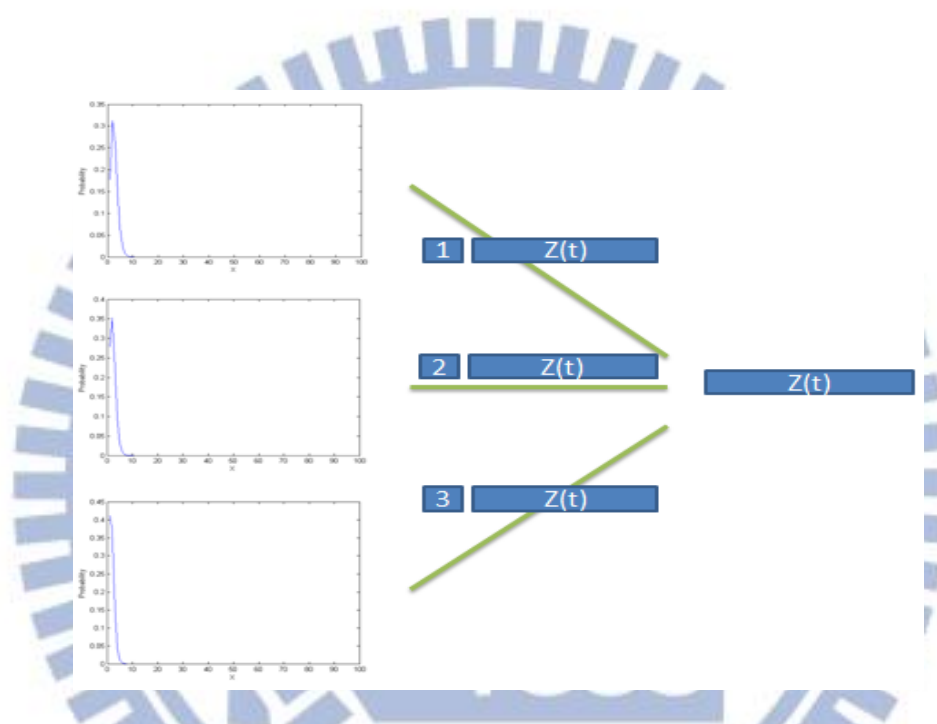


圖 4-17

接著我們比較條件在 $t+1$ 長度的 Z 歷史底下，看到的 X 統計圖有沒有差異。如果統計圖之間的差異很大，表示 t 時刻的環境 Z 仍是重要的，才會導致 X 的分佈不同，可由此知道 X 與 $t+1$ 時刻之前的 Z 仍很有關係，反之如果在 $t+1$ 長度的 Z 歷史底下，看到的 X 統計圖沒有差異，表示 $t+1$ 時刻之前的 Z 對當前的 X 已無影響。隨著 t 的增加就可以了解 X 與 Z 的歷史的關係。至於統計圖的差異度，將由以下兩個測度來決定。

4.3.1 Variance norm

在這一段落中，先以簡單的計算方式，不同的統計圖減去它們的平均統計圖，來決定統計圖間的差異，算法如下

$$\text{variance norm} = \langle \langle \sqrt{\sum_i (P_n(i) - Q(i))^2} \rangle_n \rangle_z$$

其中Q為不同統計圖的平均統計圖， $P_n(i)$ 則為第 n 個統計圖，i 為統計圖 P_n 上的第 i 位置。如 4.3 節中介紹，以 t+1 長度的某條 Z 路徑為條件，統計在此條件下看到 X 值的 histogram，將這些所有可能的 histograms 分別減去他們的平均統計圖之後平方並相加完之後再開根號，對 n 個可能的統計圖取平均之後，再對所有可能出現的 Z 路徑取平均。

取時間長度 t+1 時，觀察 X 統計圖差異，將該差異用 variance norm 表達：

t	1	2	3	4	5	6	7	8	9
Norm	0.03207	0.03039	0.02885	0.02777	0.02599	0.02441	0.02356	0.02270	0.02181

將它做成圖即以下的曲線：

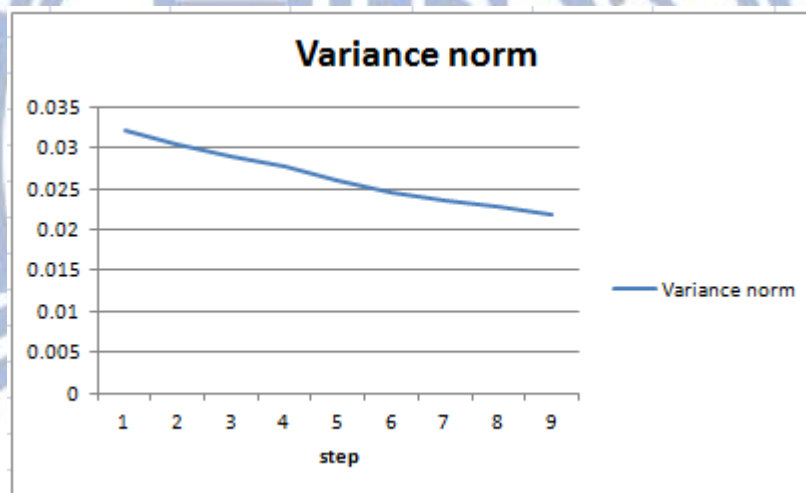


圖 4-18

由上圖可以觀察到 Norm 隨著 t 的增加越來越小，由 4.3 節中已說明若統計圖間的差異越小，代表著 protein 與越早之前 mRNA 的濃度越沒有關係。

4.3.2 KL-Divergence

這段中利用不同的分析方法_KL-Divergence，希望能與之前的 variance norm 做比較，觀察是否有一致的歷史記憶消失傾向。在此 KL-Divergence 定義為

$$D_{KL} = \langle \langle \sum_i \ln \left(\frac{P_n(i)}{Q(i)} \right) P_n(i) \rangle_n \rangle_z \quad (4.3.1)$$

其中 $Q(i)$ 為不同統計圖的平均在 i 位置的值， $P_n(i)$ 則為第 n 個統計圖在 i 位置的值。

與 4.3.1 類似的，先將 X 時刻之前 $t+1$ 時間之內的某條 Z 路徑，以不同 $t+1$ 長度的 Z 數列為條件，統計看到的 X 值的 histograms，將這些所有可能的 histograms 分別除去他們的平均統計圖之後對自己的 histogram 相乘後加總起來，對 n 個可能的統計圖取平均之後在對所有可能出現的 Z 路徑取平均。

取時間長度 $t+1$ 時，觀察 X 統計圖差異，將該差異用 KL-divergence 表達：

	1	2	3	4	5	6	7	8	9
KL_Divergence	0.00681	0.00644	0.00611	0.00593	0.00566	0.00553	0.00557	0.00550	0.00545

將表格 data 做圖即以下曲線：

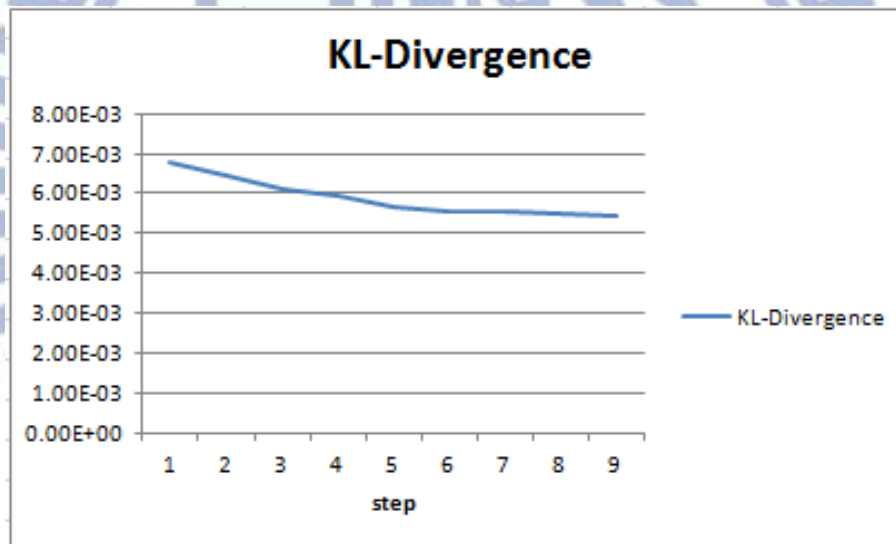


圖 4-19

比較圖 4-18 與圖 4-19，兩圖都表示都隨著歷史路徑 t 的增加，環境的歷史影響會越來越小，也就是說現在的 protein 與前一刻的 mRNA 最有關係，與越早之前 mRNA 的關係會越來越小。

4.4 當環境變動很劇烈時的 Extrinsic and Intrinsic

這裡我們嘗試將速率常數調整成：

$$\lambda_{\text{ext}} = 1500, \lambda_{\text{int}^z} = 25, \beta_{\text{ext}^z} = 400, \beta_{\text{int}^x} = 5$$

讓 mRNA 的變動相對於 protein 相當劇烈，得到下面 Extrinsic 與 Intrinsic 的變異數隨著歷史路徑長度 t 的增加的變化圖，其中綠線表示理論值，藍線與紅線代表不同 data 量下的分析，也就是分別以一次實驗中的 200 萬步與 2000 萬步，拆解成不同 step 的多次實驗來計算 Extrinsic 與 Intrinsic 的變異數：

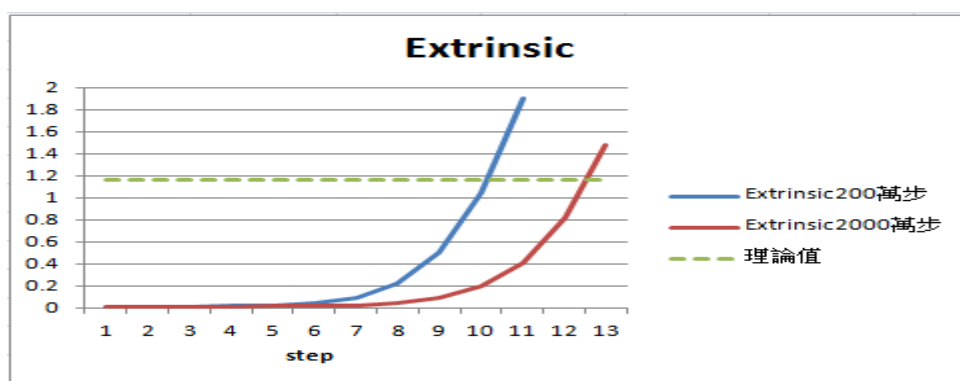


圖 4-20



圖 4-21

將圖 4-20、圖 4-21 和圖 4-12、圖 4-13 相比，Extrinsic 與 Intrinsic 變異數的差異分別隨著歷史路徑(step)增加而增大。並且不會趨向精確解，也就是利用環境的歷史無法精確的拆解這種環境變動很激烈的噪音。

4.5 Time-Correlation

我們希望借由 time-correlation 的方法來了解現在的 protein 與 T 時刻前 mRNA 的關係，

其中所使用的速率常數如下：

$$\lambda_{\text{ext}} = 15 \cdot \lambda_{\text{int}}^z = 175 \cdot \beta_{\text{ext}}^z = 4 \cdot \beta_{\text{int}}^x = 410$$

在圖 4-22，有兩條隨時間演化的路徑(A)與(B)，在 T=1 時我們將每一時刻(B)與其前一時刻的(A)分別相乘後取其平均，即代表(B)與前一時刻(A)的相關性。同樣的 T=2 則說明(B)與前 2 時刻(A)的相關性。

利用這個方法，隨著 T 的增加可以用來了解 protein 與 T 時刻前 mRNA 之間的關係。

EX:

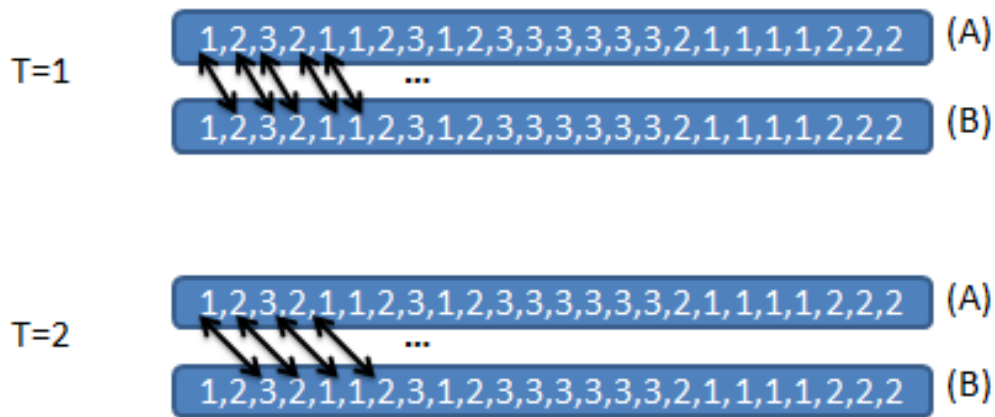


圖 4-22

4.5.1 Z 的自相關性

利用

$$\frac{1}{t} \sum_{t=1}^{\infty} (Z_t)(Z_{t+T}) \quad (4.5.1)$$

計算 time-correlation 隨著 T 的變化，得圖 4-23 曲線：

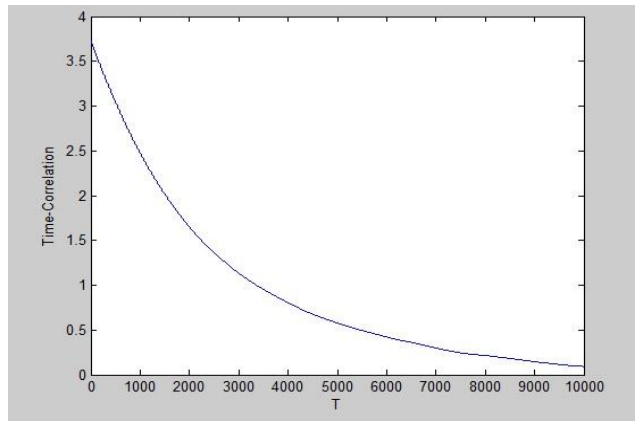


圖 4-23

由圖 4-23 得 time-correlation 在一開始的時候為最大值，接著隨著 T 的增加而越來越小，表示現在的 mRNA 與前一刻自己的狀態最為相關，離現在越遠的過去即越不相關。半衰期約在 T=1800。

4.5.2 X 的自相關性

利用

$$\frac{1}{t} \sum_{t=1}^{\infty} (X_t)(X_{t+T}) \quad (4.5.2)$$

計算 time-correlation 隨著 T 的變化，得圖 4-24 曲線：

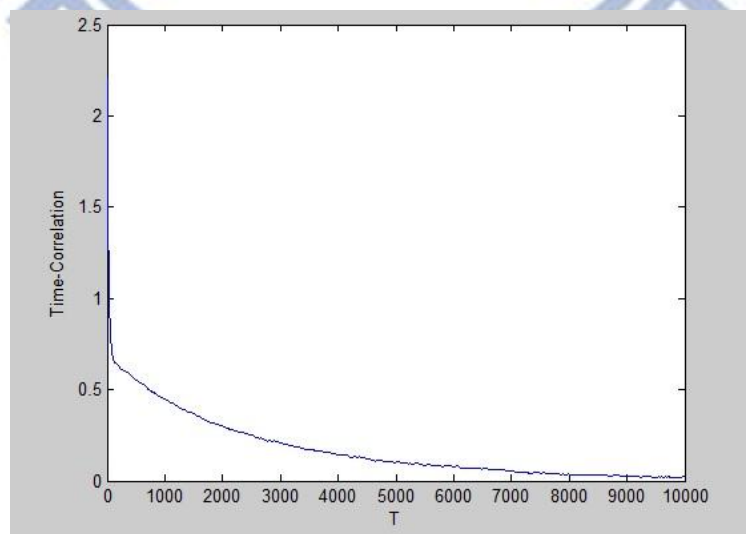


圖 4-24

由於都是在測試自相關性，在這裡與 Z 的自相關類似，由圖 4-24 得 time-correlation 在一開始的時候為最大值，接著隨著 T 的增加而越來越小，也就是 protein 與前一刻自己的狀態最為相關，離現在越遠的過去即越不相關。半衰期也約在 T=2000。

4.5.3 X 與 Z 相關性

接著利用

$$\frac{1}{t} \sum_{t=1}^{\infty} (Z_t)(X_{t+T}) \quad (4.5.3)$$

計算隨著 T 的增加，X 與 T 時刻前 Z 的 time-correlation 變化，結果畫出下圖曲線：

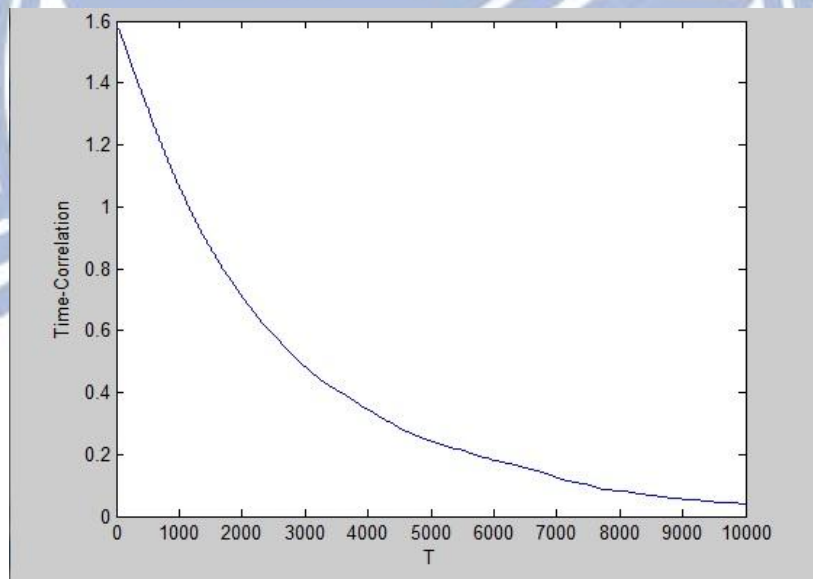


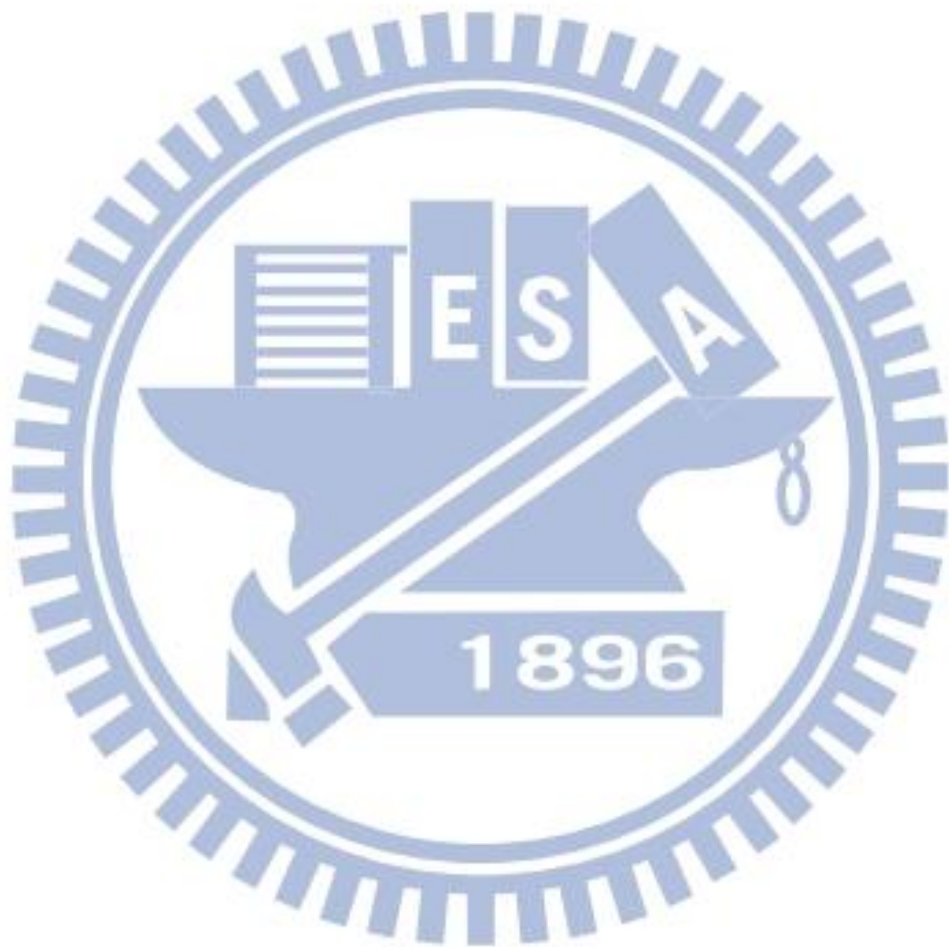
圖 4-25

如圖 4-25 可見隨著 T 的增加 correlation 大致都是越來越小，表示著 protein 與越久之前的歷史環境越不相關。

但是在 T<20 時，time-correlation 卻是越來越大，表示圖 4-16 彩色線(X)稍微左移會增加關聯性(4.5.3)，因 sum 裏頭大的(Z_t)(X_{t+T})增多。

由 4.3 節與 4.5 節的討論都可以得到相同的結果，也就是現在的 protein 與前一刻的環境

mRNA 最為相關，隨著歷史路徑的增加則 mRNA 的影響越來越小。半衰期也約為 $T=1800$ 。在此組速率常數如下 mRNA 自相關性，protein 自相關性，及 mRNA-protein 相關性的半衰期差不多。



第5章 結論與未來工作

在本文中大部分的狀況下，可以了解到 mRNA 轉錄過程中想要精確的分解內外噪聲，就必須考慮環境的所有歷史，除了某些特別的條件下($\tau_z \gg \tau_x$)，利用對當下環境為條件對噪聲做拆解就已足夠精確，並且由 4.2.4 節可知道環境的歷史在不同的速率常數下會有不同程度上的影響。

但依然有許多問題需要做進一步的探討：

1. 在 4.3 節中是比較 t+1 時間內不同統計圖的差異來了解 X 與 Z 歷史的關係。但其實也能藉由比較 t+1 時間內對 X_{t+1} 不同的統計圖與 t 時間內 Z 路徑對 X_t 平均統計圖間的差異來了解 X 與 Z 歷史的關係。
2. 4.4 節中在環境激烈變動的狀態下，利用環境的歷史為條件並不能精確的拆解內外噪聲，或許能藉由依次對影響環境的速率常數做些微的更改，來研究在環境的變動到達多劇烈的變化時，利用環境的歷史為條件才無法正確的拆解內外噪聲。
3. 在圖 4-14 與圖 4-15 中，不同的 kij 隨著歷史長度(step)的增加，Extrinsic 與 Intrinsic 變異數的改變量有不同的變化，我們猜想可藉由(2.4.1) $\langle \sigma_{X_t | Z[0,t]}^2 \rangle$ 與 $\sigma_{\langle X_t | Z[0,t] \rangle}^2$ 對 step 的微分可以獲得這些改變量，也就是當決定了 kij 時也就同時能藉由這個微分的方式算出這些改變量。
4. 在 4.5.2 節中利用 time-correlation 所做的分析尚未完全清楚為何在 $T < 100$ 時會有一個 drop，4.5.3 節中也尚未清楚會和在 $T < 20$ 時 time-correlation 的值會上升。
5. 取不同歷史長短只能做到 step=13，能研究更長的歷史將更有助於分析。

參考文獻

[1]Paulsson J (2005) Models of stochastic gene expression. *Phys Life Rev* 2:157–175

[2]Hilfinger, Andreas, and Johan Paulsson. "Separating intrinsic from extrinsic fluctuations in dynamic biological systems." *Proceedings of the National Academy of Sciences* 108.29 (2011): 12167-12172.

[3] Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.

