LaTeX 進階應用

蔡怡玲 學號:79478006

2005,10,1

1 前言

經過禮拜一課堂上, 老師說最好將每個章節先用 article 作編寫成很多個文件, 最後再將所要的文件一個個用 include 的指令引入文章之中, 而 LaTeX 就會自行幫你編排章節順序和編號了。這樣的好處是, 在編寫程式的時候才不會造成 .ctx 檔內的程式指令太過繁複且凌亂, 萬一有程式出錯比較容易找到做修改。所以, 我決定還是先用這個方法比較好, 因爲在使用 book 的時候, 遇到不少困難, 比較不容易編排。

這禮拜所教的內容有很多, 主要是在教一些 LaTeX 的進階用法。例如: 編寫各式的表格, 以及如何將圖檔引入、表格和數學式子做編號, 並且在文章中引用圖表做說明的方法。有兩個很重要的指令 \label 和 \reg 是要引入文章中所用到的。另外, 在編寫表格時還有一些 LaTeX 裡面內建的巨集套件可以使用。所以, 可以利用網路上 LaTeX 的使用手册來幫助你做更精美的編輯喔!

禮拜一看到大家的作業之後, 哎呀!大家都好厲害喔!而且很用心在做作業。雖然自己也不是隨便做一做 (我我我... 我也是花了好久的時間研究耶!), 但是同學的作業還有很多地方是值得我好好學習的地方。這禮拜我想把一些常用到的機率分配, 以及數學式子做簡單的介紹。其實順便可以當成我之後做報告的程式檔案, 以後如果要編寫數學式子, 或是表格都可以直接找到, 不需要重新鍵入。當然是要利用 LaTeX 裡面提供的製作圖表方法, 以及數學式子編寫方法, 還有引入一些精美圖檔來做介紹摟。

2 常用的機率分配

隨機變數 (random variable) 是隨機實驗中對應樣本點的實數值函數, 指其變量 x 的發生是隨機的。而所謂隨機 (random) 是指各變量的發生隨著某一「機率」而發生。隨機變數其數值的特性分爲間斷隨機變數 (discrete random variable) 與連續隨機變數 (continuous random variable)。

2.1 間斷隨機變數

間斷隨機變數(discrete random variable) 變量 X 的個數是有限的, 或個數無限但爲可數的, 我們稱爲間斷隨機變數或不連續隨機變數。間斷隨機變數是探討各個變量發生機率的分配情形, 通常以機率函數、期望值、變異數, 表現其特質。

間斷隨機變數的機率密度函數的特性:

設 X 爲間斷隨機變數, 其便量爲$X_1, X_2, ...X_n$, 對應的每一個數值有唯一的機率與之對應, 該機率值可以表示爲$f(X = x_i)$或$f(x_i)$, 並滿足下列條件:
1. $0 \leq f(X_i) \leq 1$
2. $\sum_{i=1}^{n} f(X_i) = 1$
則$f(X_i)$爲 X 的機率質量函數 (probability mass function), 簡稱 pmf

下面表1我列出幾個常用的離散型分配, 將他用表格的方式整理起來, 呈現給大家。

表 1: 離散分配

分配	機率函數	期望值	變異數
$Ber(p)$	$P(X=x\|p)=p^x(1-p)^{n-x};$ $x=0,1;\quad 0\leq p\leq 1$	p	$p(1\text{-}p)$
$Bin(n,p)$	$P(X=x\|n,p)=\binom{n}{x}p^x(1-p)^{n-x};$ $x=0,1,..,n$	np	$np(1\text{-}p)$
$U(1,n)$	$P(X=x\|N)=\frac{1}{N};x=1,2,..,N;$ $N=1,2,..$	$\frac{n+1}{2}$	$\frac{(n+1)(n-1)}{12}$
$Hyp(N,K,n)$	$P(X=x\|N,M,K)=\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$ $x=0,1,..K;M-(N-K)\leq x\leq K$	$\frac{nK}{N}$	$\frac{(N-n)}{N-1}\frac{nK}{N}(1-\frac{K}{N})$
$NB(r,p)$	$P(X=x\|r,p)=\binom{r+x-1}{x}p^r(1-p)^x$ $x=0,1,...;\quad 0\leq p\leq 1$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
$Po(\lambda)$	$P(X=x\|\lambda)=\frac{e^{-\lambda}\lambda^x}{x!}$ $x=0,1,...;\quad 0\leq\lambda<\infty$	λ	λ

表1是讀過統計學的同學都應該曾經遇到的一些離散型分配, 我們特別觀察第二欄 x 的機率值函數之間的不同, 另外注意 X 變量的數值 x 都爲可數的有限或是無限的。再來, 我們可以發現每個分配都有屬於自己的母數 (參數), 要特別注意參數彼此之間的不同, 以及個別對分配的影響。

2.2 連續隨機變數

連續隨機變數(continuous random variable) 的變量 X 是不可數的, 其數值介於某

一個區間, 而在此區間內的數值是無限的, 不可數的。其樣本空間的樣本點是密接的, 機率是以一個區間的面積來表示, 我們稱爲連續隨機變數。通常以機率函數、期望值、變異數、分配偏度和峰度來表現其特質。
連續隨機變數的機率密度函數的定義:

設 X 爲連續隨機變數, 其值爲 $a \leq X \leq b$, 若$f(X)$滿足下列條件:
1. $f(X) \geq 0$
2. $\int_a^b f(X)dx = 1$
則$f(X)$爲 X 的機率密度函數 (probability density function), 簡稱 pdf

下面表2我列出幾個常用的離散型分配, 用表格的方式整理起來, 呈現給大家。

表 2: 連續分配

分配	機率函數	期望值	變異數
$Beta(a,b)$	$f(x\|\alpha,\beta) = \frac{1}{\mathbf{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ $0 \leq x \leq 1;\quad \alpha > 0,\quad \beta > 0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
$Gamma(\alpha,\beta)$	$f(x\|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}$ $0 \leq x < \infty$	$\alpha\beta$	$\alpha\beta^2$
$Norm(\mu,\sigma^2)$	$f(x\|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < \infty$	μ	σ^2
$U(a,b)$	$f(x\|a,b) = \frac{1}{b-a}$ $a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Exp(\beta)$	$f(x\|\beta) = \frac{1}{\beta}e^{-\frac{x}{\beta}}$ $0 \leq x < \infty,\quad \beta > 0$	β	β^2

由表2我們不難看出連續型機率函數通常都很複雜, 而且變數的範圍也都大不相同, 因

此熟讀這個機率函數表是很重要的。另外和離散型分配依樣, 每個分配都有自己的母數 (參數), 有的只有一個, 有的不只一個, 但是每個參數對於分配來說都是扮演很重要的部分, 也要特別注意。

3 常用的連續分配

下面的幾個小節, 我想介紹幾個在統計上很常用且很重要的連續型機率分配。這些分配各有一些特色, 值得我們進一步做研究。他們在統計推論上也扮演著重要的角色。

3.1 常態分配

常態分配(normal distribution), 有時稱爲**高斯分配**(Gauss disturbution), 其圖形 (即次數分配曲線) 呈鐘形, 稱爲常態曲線 (normal curve)。常態分配有兩個參數, 通常記成μ跟σ^2, 分別爲其平均數及變異數。有平均數μ跟變異數σ^2(通常寫成n(μ,σ^2))。常態分配的pdf 爲

$$f(x|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty \tag{1}$$

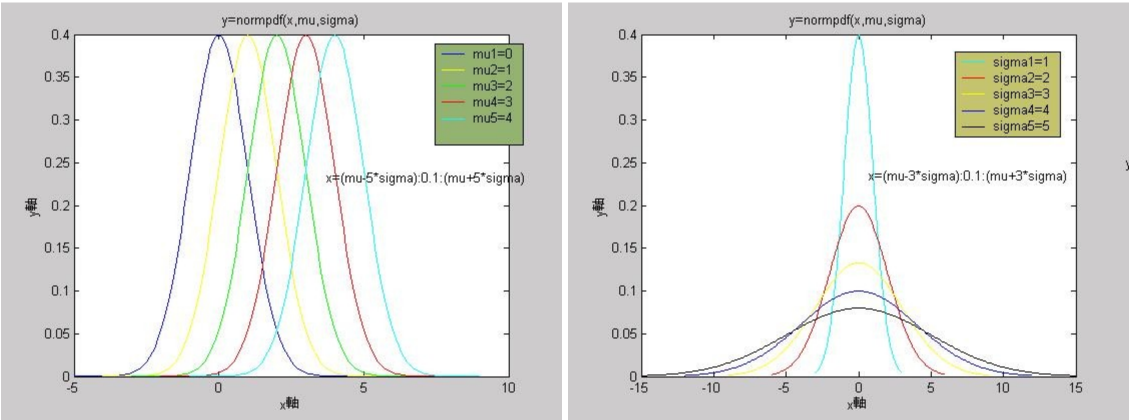
常態分配的 mean,variance,mgf 爲

$$\mathrm{E(X)}=\mu \quad \mathrm{Var(X)}=\sigma^2 \quad \mathrm{M}_X(t)=e^{\mu t+\frac{\sigma^2 t^2}{2}}$$

常態分配的特性:

- 單峰、對稱於$x=\mu$、鐘型連續的次數分配。
- 對於所有的 X 值而言,Y 值永遠爲正。
- 常態曲線與 X 軸間的全部面積爲1, 部分曲線下的面積可視爲相對次數。

常態分配的重要性之一在他的兩個參數,μ(平均數) 與σ(變異數), 提供分配確切的形狀及位置的完整資料。這個特性, 以μ及σ爲決定因素, 不只限定在常態 pdf, 而是共用於位置-尺度族 (location-scale family)。利用下面圖 1 我們就來觀察兩個參數μ及σ對於分配的影響。



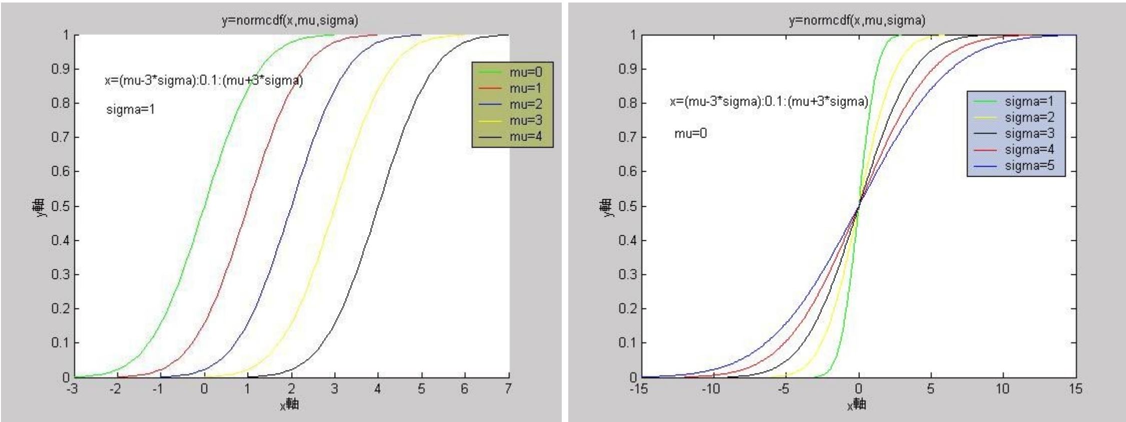
(a) (b)

圖 1: 常態分配pdf 圖形 Y = normpdf(X,MU,SIGMA)

- 圖 1 (a) 爲固定$\sigma = 1$改變μ, 我們可以觀察到μ改變並不會影響整常態分配鍾型對稱的圖形, 但其圖形中心位置會平移。因此,μ我們稱爲位置參數。
- 圖 1 (b) 爲固定$\mu = 0$改變σ, 我們可以發現σ值越大圖形呈現越平緩的分布, 但仍然爲對稱鐘型的常態分配圖形。而隨著σ值越大, 線條越趨於平緩, 是因σ爲母體變異數, 變異數大時, 圖形會分布較爲分散。因此,σ我們稱爲尺度參數。

下面我們來觀察他們的 cdf 圖形

圖 2 (a) 爲圖 1 (a)固定$\sigma = 1$改變μ的 cdf 圖形。圖 2 (b) 爲圖 1 (b)固定$\mu = 0$改變σ的 cdf 圖形。由 (b) 圖可以明顯的發現當σ越大時,pdf 圖—圖 1 (b) 分布越尖峰, 而



(a) (b)

圖 2: 常態分配cdf 圖形 Y = normcdf(X,MU,SIGMA)

cdf 圖—圖 2 (b) 累積的速度也越快, 所以越陡。

另外, 若$X \sim n(\mu, \sigma^2)$, 則隨機變數$Z = (\frac{X-\mu}{\sigma})$有個 n(0,1) 分配, 也就是標準常態分配。常態 pdf 在$x = \mu$有極大值及在$x = \mu \pm \sigma$有反曲點。而在平均數的1、2、3倍標準差內的機率是

$$P(|X - \mu| \leq \sigma) = P(|Z| \leq 1) = 0.6826$$
$$P(|X - \mu| \leq 2\sigma) = P(|Z| \leq 2) = 0.9544$$
$$P(|X - \mu| \leq 3\sigma) = P(|Z| \leq 3) = 0.9974$$

此數字值可以由許多電腦軟體或圖表中取得, 通常被討論的兩位數分別爲0.68,0.95及0.99, 雖然他們不代表四捨五入值, 但他們是被常用的值。

在許多常態分配的應用中, 有一項重要的應用, 即常態分配作爲其他分配的近似分配 (譬如: 中央極限定理即爲一例)。如: 若$X \sim$ binomail(n,p), 則 EX=np,VarX=np(1-p), 在適當條件下,X 的分配可以近似於常態分配 n(np,np(1-p))。這裡所稱「適當的條件」是 n 要大而 p 不能太靠近0或1。「n 要大」是要使 X 的離散值足夠多, 以致合理近似連續分配; 而 p 不能太靠近0或1, 即「p 接近中間」, 是要使二項值接近如常態的對稱性。二項分配近似常態的一般使用規則是在$np \geq 5$且$n(1-p) \geq 5$的情形。

3.2 T分配

若$Z \to N(0,1)$, $U \to \chi^2(v)$, 且$Z \amalg U$,

定義 $T = \dfrac{Z}{\sqrt{\dfrac{U}{v}}} \to t(v)$ 則稱 T 服從自由度為 v 之 t 分配, 記為$T \to t(v)$。我

們可以透過二維變數變換得到 T 分配 pdf 為

$$h(t) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\Pi}}(1+\frac{t^2}{v})^{-\frac{v+1}{2}}, -\infty < t < \infty \tag{2}$$

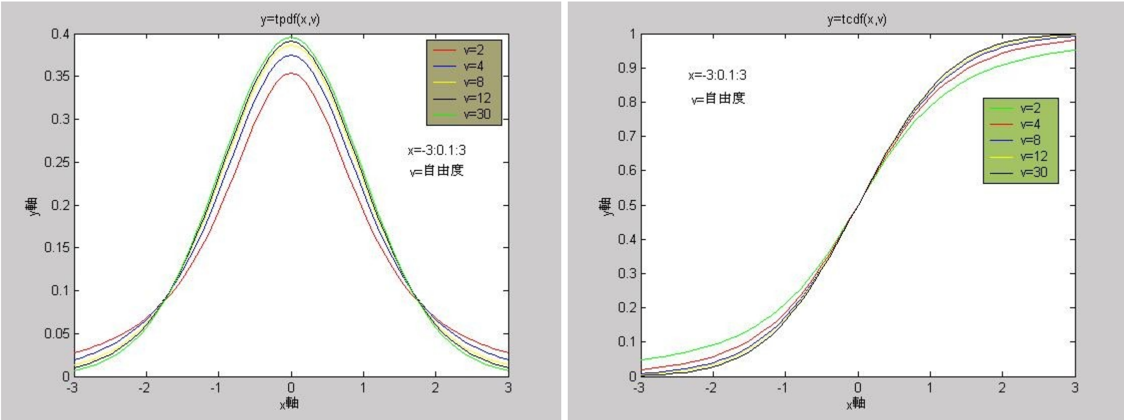
t分配具有下列特性:

1. 若 v=1, 則$h(t) = \dfrac{1}{\Pi(1+t^2)}$,$-\infty < t < \infty$, 特稱此分配為柯西分配 (Cauchy Distribution)。
2. t 分配為一對稱於0的分配 (即$f(t) = f(-t)$)。
3. $E(T) = 0(v > 1)$,$Var(T) = \frac{v}{v-2}(v > 2)$

T分配圖形與常態分配圖形相似, 都具有對稱於零、單峰及鐘形的特性。若$x_1, x_2, ...x_n$為取自常態母體 N(μ,σ^2) 的隨機樣本, 且當母體σ^2未知時, 我們採用t統計量

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \tag{3}$$

下面我們來看圖3, 觀察 t 分配的分布圖形。



(a) (b)

圖 3: T分配 pdf 圖形 Y = tpdf(x,v)

由圖3觀察, 我們知道 T 分配圖形的散佈比常態分配圖形大, 因其樣本較小分散程度較大, 則 T 分配圖形的尾端具有較多的機率, 尾端較厚。t 分配自由度越大其圖形越接近常態, 因自由度爲 (n-1), 自由度越大樣本數越大, 則變異較小。

3.3 卡方分配

若 $Z_1, Z_2, ...Z_v \overset{i.i.d}{\rightarrow} N(0,1) \quad \Rightarrow \quad \sum_{i=1}^{v} Z_i^2 \rightarrow \quad \Gamma(\alpha = \frac{v}{2}, \lambda = \frac{1}{2})$

卡方分配的pdf 爲:

$$f(x|v) = \frac{1}{\Gamma(\frac{v}{2})2^{\frac{v}{2}}} x^{\frac{v}{2}} e^{-\frac{x}{2}}, \quad 0 \le x < \infty; \tag{4}$$

卡方分配的特性:

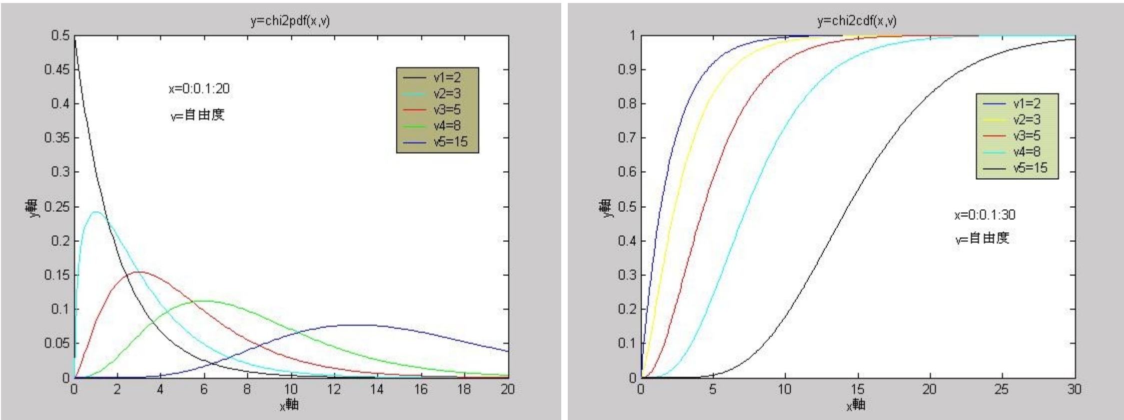
1. 通常若隨機變數 X 的機率密度函數樣本甚小時我們可以採用卡方分配;

2. 卡方值均爲正的, 其全距自0延伸至無限大;

3. 卡方分配是一 Gamma 分配$\alpha = \frac{v}{2}$,$\lambda = \frac{1}{2}$, 因此平均數皆等於其自由度v;

4. 若自由度$>$ 30, 則可採用常態的近似分配代替卡方分配。

利用下面圖4我們針對卡方分配的函數圖形做觀察。



(a) (b)

圖 4: 卡方分配pdf 圖形$Y = chipdf(x, v)$

由圖4(a) 我們知道卡方分配參數爲自由度'v', 分配圖形急劇右偏 (偏向縱軸), 但是隨著樣本數 (自由度) 的增加, 卡方分配漸趨對稱。圖形越接近對稱鍾型的常態分配, 因此通常若自由度$>$ 30, 則可採用常態的近似分配來代替卡方分配。

3.4 F分配

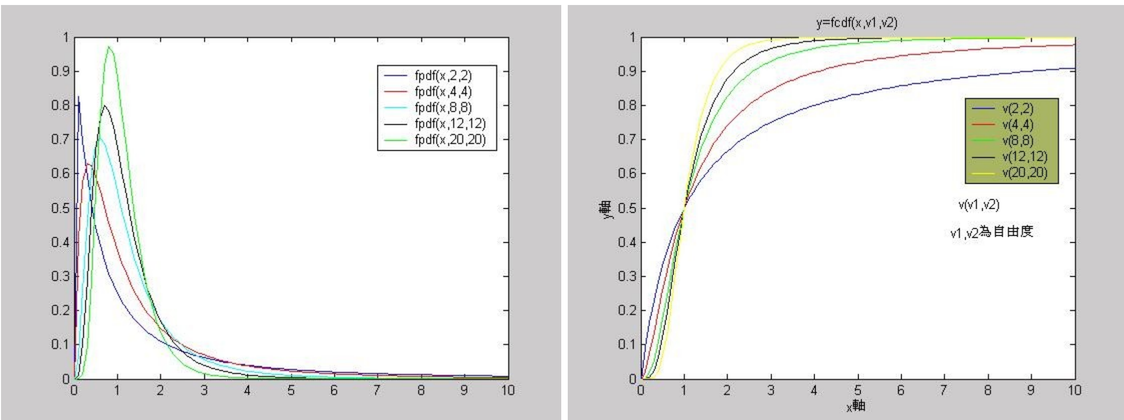
若$U \to \chi^2(v_1), V \to \chi^2(v_2)$, 且$U \,\mathrm{II}\, V$, 則定義 $F = \dfrac{\frac{U}{v_1}}{\frac{V}{v_2}} \to F(v_1, v_2)$

稱 F 服從自由度爲(v_1, v_2)的 F 分配,pdf 爲

$$f(x|v_1,v_2) = \frac{\Gamma(\frac{v_1+v_2}{2})}{\Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})} \left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}} \frac{x^{\frac{v_1-2}{2}}}{(1+(\frac{v_1}{v_2})x)^{\frac{v_1+v_2}{2}}} \tag{5}$$

F分配具有下列特性:

1. 每一個 F 值有兩個自由度。v_1代表分子的自由度,v_2而是分母的自由度
2. 當分子的自由度爲1時,F 分配即等於某一 t 分配, 其自由度與 F 分配之分母自由度相等。
3. F 統計量=(樣本變異數比)/(母體變異數比)
   =(樣本平均數之間的變異) / (在相同樣本內個體間的變異),
   用來檢定數個母體是否有相同的平均數。



(a) (b)

圖 5: F分配 pdf 圖形$Y = fpdf(x, v_1, v_2)$

由圖5可以觀察他的機率分配, 和累積機率分配。

4 心得

這禮拜的作業我花了更多的時間去編排, 因爲表格內容的編寫, 需要花很多的時間。另外, 我的表格內容很多, 所以表格很大, 相對的佔掉版面很多空間。所以當我在編寫文字敍述的時後會不容易控制版面, 因爲如果剩下頁面不夠擠下表格的時候, 它會自動跳到下一面。而數學式子和圖檔更容易發生類似情形, 所以調整數學式子以及圖表和引入文字敍述之間, 花了我很多的精力。

插入圖檔後的文字敍述和版面呈現, 是我這次作業遇到的一個不容易解決的困難。但是這禮拜的作業裡面用了許多 LaTeX 編排的套件, 我學到了不少巨集指令, 而且數學的程式大致上做了一個整理, 算是收穫不少喔!

熊熊給他發現, 現在已經是 2005.10.3 凌晨 00:30 糟了, 不妙! 又經過了一個禮拜, 我我我怎麼一點長進都沒有, 還是遲交了。sorry 阿老師, 辛苦你了。