統計碩二 連振宇 高統推期中報告 主題：復發事件分析

1. 存活分析中的復發事件

   在許多醫學研究裡，被觀察的對象可能在試驗時間裡會經歷多種同樣的事件，因此除了考慮單一事件外，多元事件必須納入考量。舉例而言，癌症腫瘤轉移所造成的多次腫瘤復發等，此種事件變稱為復發事件（recurrent event）。而分析此種資料時，可以分成針對事件時間（event times）或是間隔時間（gap times）去探討。

   a. 事件時間（event times）

      研究個體以從起始事件到發生事件的事件時間（event times）為時間尺度。但兩事件時間會有一部分交疊，因此造成若研究此種資料結構，必然會存在事件時間之間的相關性。

   b. 間隔時間（gap times）

      研究個體發生兩兩相繼事件之間時間，也就是間隔時間（gap times）為時間尺度。雖可有效避免事件時間的相關性問題，但在間隔時間不獨立下，若有設限則會有 induced informative censoring。

   c. 分析方法

      分析並建模復發時間可藉由考慮在一短時間區間內發生事件次數的機率分配並給定事件歷史（history）去分析，我們定義

      $\mathrm{N(t)}, \mathrm{t} \geq 0$ is counting process

      $\Delta N(t) = N(t+\Delta t^{-}) - N(t^{-})$ 代表在 $[\mathrm{t}, \mathrm{t}+\Delta \mathrm{t})$ 區間中的事件個數

      $H(t) = \{N(s) : 0 \leq s < t\}$ 表 process 在時間 t 前一刻的歷史（history）

      $T_1, T_2, \cdots, T_k$ 代表事件時間

      則我們可以用兩種函數去進行分析，分別是強度函數（intensity function）與速率函數（rate function），其中

      $$\lambda\big(t \big| H(t)\big) = \lim_{\Delta t \to 0} \frac{P\big(\Delta N(t) = 1 \big| H(t)\big)}{\Delta t}$$

      代表了強度函數(intensity function)，其概念類似於風險函數的概念，代表給定過去的歷史下，已知上一瞬間的狀態，在下一個瞬間發生事件的機率，且其可以唯一地建立點過程（point process）的機率結構；而

      $$\lambda(t) = \lim_{\Delta t \to 0} \frac{P\big(\Delta N(t) = 1\big)}{\Delta t}$$

      代表了速率函數(rate function)，其代表了 t 單位時間內的平均事件次數，但此函數並無法唯一地建立點過程的機率結構。

2. Poisson process in recurrent event analysis

在復發事件中我們可以用 counting process 來計算事件發生的次數(events counts)，因此，這種基於計算事件發生次數的過程，我們可以使用 Poisson process，根據其定義，其本質就是一種 counting process，且具有獨立增量，且與 Poisson distribution 的概念：計算發生事情的次數也吻合，而通常會以分析事件時間為主，並基於此去建模分析。但也並非所有實際情況都合適用此種方式，基本上需要研究個體會經常發生感興趣的事件，且發生事件本身也應獨立於 process 本身。舉例而言，氣喘發作因彼此之間較屬於偶然性，因此不會改變 process 本身的性質，某種程度而言可以想像程是具有「獨立增量」的性質，而若是前例中的癌症發生腫瘤轉移次數，由於彼此之間應存在相關性，因此不太適合，應以分析其事件相關性或是依據其間隔時間為主。

在 Poisson process 中，由於獨立增量的性質，因此可注意到

$$\lambda\big(t\big|H(t)\big)=\lim_{\Delta t\to 0}\frac{P\big(\Delta N(t)=1\big|H(t)\big)}{\Delta t}=\lim_{\Delta t\to 0}\frac{P(\Delta N(t)=1)}{\Delta t}=\lambda(t)$$

也就是其強度函數正巧等於速率函數，而由於速率函數的性質(t單位時間內的平均事件次數)，這也再次呼應了 Poisson process 應用在復發事件分析中是以計算事件次數(events counts)為主，且實際上由上述強度函數定義出來的 Poisson process 也會是個 Markov process，因其在單位時間內的機率僅與 t 有關而獨立於歷史，而這也再次呼應了 $\Delta N(t)$ 與 $H(t)$ 獨立的事實。而 Poisson process 可以在區分為同質(Homogeneous)與異質(Heterogeneous)，也就是滿足穩態條件(stationary condition)與否，但無論何者都須滿足獨立增量。

a. Homogeneous and Heterogeneous Poisson process in recurrent events data analysis

1. Poisson Regression Model

此模型是依照強度函數的概念加入共變數後建模如下

$$\lambda_i\big(t\big|H(t)\big)=\lambda_0(t)exp\big(\beta Z_i(t)\big)$$

其中的 β 為模型估計參數，$\lambda_0(t)$ 為基底風險函數，$Z(t)$ 為時間相依共變數。在此模型中，假設 λ_0 為常數，且過去發生的事件並不會影響後續事件發生的速率，邊際事件發生速率也不會隨時間而改變，而這就是 Homogeneous Poisson process 的特性。好處是模型容易建

構，但缺點正是其假設事件發生速率同質性在實務上通常不存在。

2. Marginal Failure Rate Model (Pepe and Cai (1993), Lawless and Nadeau (1995)

若定義

$$\mathrm{k}_i \stackrel{\text{def}}{=} \Lambda(t_2) - \Lambda(t_1)$$

其中Λ為累積風險函數，且

$$N_i(t_2) - N_i(t_1) \sim \mathrm{Poisson}(k_i)$$

也就是在時間區間 $[t_1, t_2]$ 中發生事件次數符合 Poisson distribution，且其變異數與平均數為累積風險時間函數在此兩時間點值的差異，因此可注意到事件發生速率的變異性來自強度函數(但也可以來自共變數)。而此時若考慮

$$\alpha\big(t\big|Z(t)\big) = \lim_{\Delta t \to 0} \frac{P\big(\Delta N(t) = 1\big|Z\big)}{\Delta t} = \alpha_0(t)exp(\beta Z)$$

此即為marginal failure time model，其是針對平均事件發生速率去建模，並且給定共變數下可對平均事件次數建模，且可注意到只受共變數影響，但並未給定歷史，但也並非假設與歷史無關，只是建模上並未考慮，此及邊際模型(marginal model)的概念。而此模型本質上是 Heterogeneous Poisson model，因為其允許邊際事件發生速率 $\alpha_0(t)$ 可隨時間改變。但實際上也可以是 Homogeneous Poisson process，因只要我們假設 $\alpha(t) = \alpha_0$ 則就等同假設了邊際事件發生速率 α_0 不隨時間改變，因此為穩態的條件。

3. Andersen and Gill Model (Andersen and Gill (1982))
此模型以事件數量增量(increment in the number of events)來建模，其也仍然是使用事件時間。模型為以下：

$$\lambda\big(t\big|H(t)\big) = \lambda_0(t)exp\big(\beta Z(t)\big)$$

此模型的好處為簡單推論，但缺點是為了此方面推論其假設了個體的事件時間之間的相關性可以通過過去的事件來解釋，這代表在給定共

變數的情況下，事件之間的時間增量在條件上是不相關的，因此並未考慮事件時間之間的相關性。當每個個體之間的事件之間的相關性是通過共變數來影響時，這是一個合適的模型，但實際層面上此假設通常過於強烈。

而此模型本質上是個 Heterogeneous Poisson process，因可注意到其類似於 marginal failure time model，因在此 AG model 也允許其邊際事件發生速率得以與隨時間變動，但也同樣可以改為考慮 Homogeneous Poisson process。

4. Frailty (Example : Gamma-Poisson process)
復發事件中可能會存在個體間變異(interindividual variation)，舉例而言，在研究施打疫苗對於某疾病的效力時，也許不同人之間對於疫苗的事應本身就有差異，亦即有些人就是對疫苗有效，有些人就是沒效，又或是我們希望能夠在模型中除了以 Kendall tau 或是 cross ratio 以外的方法去多測量出事件時間之間的相關性等等，而有此種情況的跡象之一即是

$$Var(N_i(t)) > E(N_i(t))$$

因 Poisson process 性質理應使上述為等號，但若變異數明顯較大則就逼需考慮個體間變異的情況了，此時可以考慮一潛在變數 (latent variable) 來表示每一個個體都共享此脆弱 (frailty) 效應，且假設不同個體間此脆弱效應獨立，來測量這相關性。

若我們令此脆弱變數為 η ，則此時給定共變數 z 下，Poisson 的速率函數建模就會變成

$$\lambda(t|z_i,\eta_i) = \eta_i\lambda_0(t)exp(\beta z_i)$$

若假設 $\eta \sim \text{Gam}(1,\theta)$，則此時

$$N_i(t|z_i,\eta_i) \propto \text{Poisson}(\eta_i E[N(t)])$$

上述稱為 Gamma-Poisson process，此時我們可以得到給定共變數 z 下此一 process 的機率函數為

$$P(N_i(s,t)=n|\mathrm{z_i})=\int_0^\infty \frac{\Gamma\left(n+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)}\frac{\left[\theta E[N_i(s,t)]\right]^n}{\left[1+\theta E[N_i(s,t)]\right]^{n+\frac{1}{\theta}}}$$

可以發現此會遵從一個 Negative Binomial distribution。且可注意到，當 $\theta \to 1$時會退化到 Poisson process，這也呼應了若此條件發生時，也就代表這個體間變異會趨近於不存在。

3. Renewal process in recurrent event.

在復發事件中我們可以用基於間隔時間 (gap times) 的方法去建構模型分析，而當感興趣的事件相對不頻繁發生，且事件發生後類似於更新狀態而可能有所改變 process，又或是對下一個事件時間的預測感興趣時，則適合使用間隔時間來分析，舉例而言可能是預測前例的癌症腫瘤轉移，又或是精神病患重複經歷住院的情況。而基於此分析，會使用更新過程 (renewal process)來進行，正因為我們分析的是間隔時間，每發生一次事件後，間隔時間就會歸零重新開始，而這也對應到了更新過程的概念。

在此情況中，先前的強度函數會變為分析每個間隔時間中的風險函數，也就是

$$\lambda\left(t\middle|H(t)\right)=h(t-T_{N(t^-)})$$

其中 h 代表兩事件之間的間隔時間之風險函數。但也需注意的是，若我們令

$$W_j=T_j-T_{j-1}$$

間隔時間的條件分布為

$$P\left(W_j>w\middle|T_{j-1}=t_{j-1},H\left(t_{j-1}\right)\right)=exp\left(-\int_{t_{j-1}}^{t_{j-1}+w}\lambda\left(\mu\middle|H(\mu)\right)du\right)$$

因此可以注意到，間隔時間原則上是不獨立的。但是在 Homogeneous Poisson process 中，由於強度函數為常數，因此在此狀態間隔時間彼此獨立且同分配，其分配可推導出其遵從指數分配，參數與強度函數有關。

值得一提的是，由於 Renewal process 有假設彼此須為獨立且同分配的條件，正如同若我們考慮精神病患出院再住院的資料型態，出院住院的間隔時間彼此之前應存在相關性，又或是先前所提的 induce informative censoring 問題，因此這些間隔時間可能會互相影響，甚至於會不同分配。此時可以考慮 modulated renewal process，其放寬了需獨立同分配的假設，並添加了 longitudinal 強度函數去調整了事件發生速率的期望值。其中 Multiplicative model 變是基於此想法去建模，其考慮

$$\lambda\big(t\big|H(t)\big) = h_0(w)exp(z(t)\beta)$$

其中

$$w = t - T_{N(t^-)}$$

為 elapsed time，也就是距離上一事件時間發生後經過多久的時間，而這裡的時間相依共變數包含了歷史，在此模型下 $W_j = T_j - T_{j-1}$ 的風險函數給定共變數下會變成

$$h_j(w) = h_0(w)exp\big(z\big(t_{j-1} + w\big)\beta\big)$$

同時我們也可以仿造在 Poisson process 中所提到的考慮脆弱變數 η，此時就會變成 conditional multiplicative model：

$$h(w|\eta_i, z_i) = \eta_i h_0(w)exp(z_i\beta)$$

其中的 η_i 同樣是假設為獨立且同分配的隨機變數。

4. 總結

Poisson process 與 Renewal process 在復發事件中用到的非常頻繁，還有許多的內容可以討論，諸如 likelihood 建構又或是 renewal function 跟 arrival time 的彼此關係等，又或是深入去討論脆弱性變數的狀態下這些 process 的性質。且也不單就是這兩種 process 或是復發事件，counting process 在單一事件分析中也佔有極為重要的角色，因為可以去衍伸出重要的 martingale central limit theorem 去討論存活分析中各種估計式大樣本下漸進的分配，而這些都非常依賴隨機過程的學習，未來若有機會希望可以整理出在存活分析中各種估計式(Product limit, Nelson-Aalen, Cox Ph model 等)在大樣本下漸進的性質，並深入去探討隨機過程的性質等。

5. 參考書籍

1. Andersen, P.K. and Gill, R.D. (1982). Cox’s regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–1120.
2. J. F. Lawless (1987) Regression Methods for Poisson Process Data *Journal of the American Statistical Association* Vol. 82,. 808-815.
3. Lawless, J.F. and Nadeau, J.C. (1995). Nonparametric estimation of cumulative mean functions for recurrent events *Technometrics* **37**, 158–168.
4. Pepe, M.S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times. J. *Am. Statist. Assoc.* **88**, 811–820.
5. Richard J. CookJerald F. Lawless (2007) *The Statistical Analysis of Recurrent Events* Springer: New York, NY.
6. Sheldon M. Ross (1996) *Stochastic Processes* Wiley.