和光同塵(下)

黃文璋

4. 無記憶性質

Cundy(1966) 說他曾自收音機中獲知, 知更鳥 (robin) 平均可再活 1.2 年, 而與牠現在年齡無關。假設某日你與一朋友相約至海邊垂釣。朋友的釣魚技術與你相彷。你抵達約定地點時, 朋友已先到了 30 分鐘, 但尚未釣上魚。你是否會覺得朋友較你有更大的機會釣上第一條? 很可能不會。原因是: 魚應不至於同情朋友已枯坐了 30 分鐘, 而讓他先釣上一條。此正如丟一銅板, 直至出現一正面才停止。如果連丟十次皆得反面, 那就是白丟了, 以後的發展就像重新丟一樣, 而不會是因已丟那麼多次, 所以正面快出現了。

我們常說歲月催人老。但如前述釣魚的例子, 有些事物, 歲月不會使其衰老, 此物還會 "活" 多久, 與一新生代相彷。此物之 "死亡", 似乎都是因爲某一突然的事件發生, 而非逐漸衰退。亦即若此物已活了 a 單位的時間, 則會再活至少 b 單位的時間之可能性, 與 a 無關。此物彷彿會忘記它自己已活了多久。此性質便稱爲無記憶性質 (memoryless property)。

年輕人通常不希望壽命有無記憶性質, 老年人一般而言, 就很希望壽命有無記憶性質。不過在戰場上, 對那些在第一線的士兵, 到底還能活多久, 很可能就有無記憶性質了。

正式地說, 一隨機變數 X, 稱爲在某一實數的子集合 S 中有無記憶性質, 若滿足對 $\forall a, b \in S$,

$$P(X > a+b \mid X > a) = P(X > b)。 \tag{1}$$

例1: 連續丟一出現正面機率爲 $p > 0$ 之銅板, 直至得到一正面才停止, 令 X 表總共的投擲數。則 X 有自 1 開始, 且參數爲 p 之幾何分佈 (geometric distribution), 以 $Ge(p)$ 表之, 即

$$P(X = k) = p(1-p)^{k-1},\ k = 1, 2, 3, \ldots。 \tag{2}$$

則

$$P(X > k) = \sum_{i=k+1}^{\infty} p(1-p)^{i-1} = (1-p)^k。 \tag{3}$$

因此

$$P(X > a+b \mid X > a) = \frac{P(X > a+b)}{P(X > a)}$$

$$
\begin{aligned}
&= \frac{(1-p)^{a+b}}{(1-p)^a} \\
&= P(X > b)。
\end{aligned}
$$

即若取 $S = \{1, 2, 3, \ldots\}$, 則對 $\forall a, b \in S$, (4) 式成立。因此自 1 開始之幾何分佈, 有無記憶性質。

有位婦人很想生個女兒, 她已連生 7 個兒子, 朋友都鼓勵她再生, 因爲那有運氣那麼壞的? 你現在知道鼓勵她再生的人是沒有道理的, 因若令 Y 表得到一女兒前, 所生之兒子數, 則 Y 有自 1 開始且參數爲 1/2 之幾何分佈。由於 Y 有無記憶性質, 所以 $P(Y > 7 + b \mid Y > 7) = P(Y > b)$, 因此那已生的 7 個兒子, 並無助於使她更快獲得一女兒。

例2: 指數分佈 (exponential distribution) 亦有無記憶性質。設 Y 有參數 λ 之指數分佈, 以 $\mathcal{E}(\lambda)$ 表之, 即 Y 之機率密度函數 (probability density function, 簡稱 p.d.f.) 爲

$$f(x) = \lambda e^{-\lambda x},\ x > 0。 \tag{4}$$

則因

$$
\begin{aligned}
P(Y > x) &= \int_x^{\infty} \lambda e^{-\lambda u} du \\
&= e^{-\lambda x}, \quad x > 0。
\end{aligned} \tag{5}
$$

故

$$
\begin{aligned}
P(Y > a + b \mid Y > a) &= \frac{P(Y > a + b)}{P(Y > a)} \\
&= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} \\
&= e^{-\lambda b},\ \forall a, b > 0。
\end{aligned} \tag{6}
$$

基本上, 幾何分佈及指數分佈爲僅有的兩個具有無記憶性質之分佈。證明見黃文璋 (1995) 第一章第 6 節。我們發現此二分佈之存活函數 (survival function) 皆爲指數函數。在此對一隨機變數 X,

$$\overline{F}(x) = 1 - F(x) = P(X > x),\ x \in R,$$

稱爲其存活函數, 其中 $F(x) = P(X \leq x)$ 爲 X 之分佈函數 (distribution function)。存活函數又稱尾部機率函數 (tail probability function)。若 X 表某物之壽命, 則存活函數 $\overline{F}(x)$ 給出此物會存活至少 x 單位時間之機率。在諸如機器可靠度 (reliability) 的探討, 醫學上某種疾病之壽命的探討, 精算學的 (actuarial) 探討, 通常都是對存活函數較有興趣。

指數函數且底小於 1 (幾何分佈底爲 $(1-p) < 1$, 指數分佈底爲 $e^{-\lambda} < 1$, 分別見 (6) 式及 (9) 式), 表 $k \to \infty$ (或 $x \to \infty$ 時), $P(X > k)$ (或 $P(Y > x)$) 趨近至 0 之速度很快。換句話說, 隨著 k 之增大 (或 x 之增大), X 要大於 k (或 $Y > x$) 愈來愈不容易, 其難度, 我們之前以指數函數的威力來形容。但壽命之存活函數若爲指數函數, 則壽命便有無記憶性質, 能再活多久, 不受現有年齡的影響。此物彷彿隨時保持一全新狀態。這眞是一奇特的現象。起初由其存活函數趨近至 0 的速度很大, 我們會以爲此物不易活很久, 沒想到卻隨時像新的一樣。

國內數學界有位資深且爲人敬重的前輩, 一向語多詼諧。他個子不高, 嘗言“身高要看從那裡量, 從天花板量起我最高”。頭頂距天花板的距離, 的確他最遠。年齡亦是如此,

一向我們是量距出生時之歲月長。但距死亡日還多久也是一不錯的比法: 距死亡日較久的便算較年輕。壽命若有幾何分佈或指數分佈, 依此新算法, 便永遠像是松柏長青。要知駐顏有術不如青春永駐, 這是恭維人時不可不留意的。

5. 波松過程

假設在時間 $t=0$ 裝了一新燈泡, 用了 X_1 的時間燈泡壞了, 立即換新 (更換的時間假設可忽略), 再用了 X_2 的時間, 第二個燈泡又壞了, 於是立即更新, 如此繼續進行下去。若燈泡的壽命 $X_1, X_2, \ldots$ 假設爲獨立且有共同分佈 (independent and indentically distributed, 簡寫爲 i.i.d.), 且令 $A(t)$ 表至時間 t 所共更換的燈泡數 (在 $t=0$ 裝的那一個不算, 所以 $A(0)=0$)。則 $\{A(t),\, t\geq 0\}$ 便稱爲一更新過程 (renewal process)。

更新過程的例子很多, 只要是依序觀察某特定事件, 事件發生之間距設爲 i.i.d., 則至時間 t 共發生的次數 $A(t)$, $t\geq 0$, 便構成一更新過程。

若事件發生之間距 X_i, $i\geq 1$, 以指數爲其共同分佈, 即設存在一 $\lambda>0$, 使 $P(X_i\leq x)=1-e^{-\lambda x}$, $x>0$, 則 $\{A(t), t\geq 0\}$ 便稱爲一參數爲 λ 之波松過程 (Poisson process)。波松過程爲一點過程 (point process, 如果發生一事件, 便在時軸上標一點, 在某時區中發生幾個事件, 便是在該時區中有幾個點), 也是一種計數過程 (counting process, 計算至時間 t 共發生幾個事件)。

爲何以波松對此過程命名呢?

令 $S_n=X_1+X_2+\cdots+X_n$, $n\geq 1$, 表第 n 個事件之發生時刻, $S_0=0$, 若 X_i 以 $\mathcal{E}(\lambda)$ 爲其共同分佈時, 如果你機率論學得夠好, 當知 S_n 有 $\Gamma(n,1/\lambda)$ 分佈。否則由求 S_n 之拉普拉斯轉換 (Laplace transformation)

$$\begin{aligned} E(e^{-tS_n}) &= E(e^{-t(X_1+X_2+\cdots+X_n)}) \\ &= (E(e^{-tX_1}))^n \\ &= (\frac{\lambda}{\lambda+t})^n = (1+\frac{t}{\lambda})^{-n}, \end{aligned}$$

亦得 S_n 確有 $\Gamma(n,1/\lambda)$ 分佈。即 S_n 之 p.d.f. 爲

$$\begin{aligned} f(x) &= \frac{\lambda^n x^{n-1}e^{-\lambda x}}{\Gamma(n)} \\ &= \frac{\lambda^n x^{n-1}e^{-\lambda x}}{(n-1)!},\ x>0。 \end{aligned} \tag{7}$$

反覆利用分部積分 (integration by parts) 可得

$$\begin{aligned} P(S_n>t) &= \int_t^\infty f(x)dx \\ &= \sum_{i=0}^{n-1}\frac{(\lambda t)^i e^{-\lambda t}}{i!},\ t>0。 \end{aligned} \tag{8}$$

現因事件 $[A(t)=n]$ 與 $[S_n\leq t<S_n]$ 等價, 故

$$\begin{aligned} &P(A(t)=n) \\ &= P(S_n\leq t<S_{n+1}) \qquad (9) \\ &= P(S_{n+1}>t)-P(S_n>t) \\ &= \frac{(\lambda t)^n e^{-\lambda t}}{n!},\ \forall t\geq 0,\ n=0,1,2,\ldots。 \end{aligned}$$

即得證對一參數爲 λ 之波松過程, 至時間 t 發生的個數 $A(t)$, 有參數 λt 之波松分佈。這是此過程以波松命名的原因。

對任二 $t, s > 0$, $A(t+s) - A(t)$ 表波松過程在區間 $(t, t+s]$ 中所發生的事件個數。可以證明 $A(t+s) - A(t)$ 有參數 λs 之波松分佈。也就是說 $A(t+s) - A(t)$ 之分佈只與區間長度 s 有關, 而與其位置無關。而且 $A(t+s) - A(t)$ 還與 $A(t)$ 獨立呢! 甚至對任意 $n \geq 2$, 及 $0 \leq t_0 < t_1 < \cdots < t_n$, $A(t_n) - A(t_{n-1})$, $A(t_{n-1}) - A(t_{n-2}), \cdots, A(t_2) - A(t_1)$, $A(t_1)$, 此 n 個隨機變數相互獨立。即在不相交區間 $[0, t_1], [t_1, t_2], \ldots, [t_{n-1}, t_n]$ 中之事件發生個數相互獨立, 這是波松過程一重要的性質, 稱爲獨立增量 (independent increments) 性質。

在很多實際的情況中, 如釣魚、交通事故、意外事件及至醫院看病等, 由於至下一事件的等待時間往往有無記憶性質, 因此二事件的間距便有指數分佈, 從而至時間 t 發生的事件數 (選一起始點定爲時間 0), 便形成一波松過程。

我們也可以另一方式來說明爲何波松過程處處可見。指數分佈的無記憶性質導致在任一小區間 $[t, t+h]$ 中, 會發生一事件之機率爲

$$\begin{aligned} P(X \leq h) &= 1 - e^{-\lambda h} \\ &= 1 - (1 - \lambda h + o(h)) \\ &= \lambda h + o(h),\ h \to 0, \end{aligned}$$

其中 X 表一有 $\mathcal{E}(\lambda)$ 分佈之隨機變數, 此處用到當 h 很小時

$$e^{-\lambda h} = 1 - \lambda h + o(h),$$

又 $o(h)$ 爲某一 h 的函數 (o 讀爲 little-oh), 且滿足

$$\lim_{h \to 0} \frac{o(h)}{h} = 0。$$

換句話說, 當 h 很小時, $o(h)$ 與 h 相比更小。很多點過程有這種性質: 在一小區間發生一個點 (或說一事件) 的機率約與此區間長成正比 (除了一更小的誤差), 而與此區間的位置無關。至於會發生兩個以上的點之機率則是極微小, 以 $o(h)$表之。注意 $o(h)$ 並非指一特定的函數, 諸如 h^2, $h^{3/2} + h^2$ 皆爲 $o(h)$, $h \to 0$。直觀上來看, 在一區間 $[0, t]$ 中會發生幾個點便有波松分佈 (利用二項分佈趨近至波松分佈的結果), 且在不相交區間中各發生幾個點爲相互獨立。波松過程便產生了。

無記憶性質, 事件發生的均勻性 (發生機率只與區間長度有關, 且區間較短時, “差不多”是線性的), 在不少實際的現象裡, 均具有此二種特性, 因此使得波松過程成爲一出現極頻繁的計數過程。

關於波松過程的討論, 以及更一般的波松過程之定義, 可參考黃文璋 (1995) 第五章。

6. 等待時間之詭論

設有一參數爲λ之波松過程$\{A(t),\ t \geq 0\}$, 譬如說, $A(t)$ 表至時間 t 共換了幾個

燈泡。則在時間 t 正在使用的那一燈泡還可用多久呢? 由指數分佈之無記憶性質, 我們知道該燈泡之剩餘壽命 (residual life 或說 remaining life), 仍爲有參數 λ 之指數分佈, 與一般正常的間距一樣。至於該燈泡已經用多久了呢? 明顯地, 其已經用的時間 (稱做現在壽命 (current life)), 不能超過 t。若以 δ_t 表現在壽命, 則易見只要 $x<t$, 則 $\delta_t>x$, 若且唯若在區間 $(t-x,t]$ 中, 沒有換燈泡, 而此機率爲 $1-e^{-\lambda x}$。故

$$P(\delta_t\leq x)=\begin{cases}1-e^{-\lambda x}, & 0\leq x<t,\\ 1, & x\geq t。\end{cases} \tag{10}$$

當 $x\geq t$ 時, 由於必有 $\delta_t\leq t$, 故 $P(\delta_t\leq x)=1$。一般我們用 γ_t 表在時間 t 正在使用的那一燈泡之剩餘壽命。

於時間 t 正在用的那一燈泡之總壽命 $\beta_t=\gamma_t+\delta_t$, 其期望值爲

$$\begin{aligned}E(\beta_t)&=E(\gamma_t)+E(\delta_t)\\ &=\frac{1}{\lambda}+\int_0^t x\lambda e^{-\lambda x}dx+tP(\delta_t=t)\\ &=\frac{1}{\lambda}+\frac{1}{\lambda}(1-e^{-\lambda t})。\end{aligned}$$

不少人都有下述經驗: 等公車時, 別人的車子總是先到, 自己要搭的卻老不來, 站牌上明明寫 15 分鐘一班, 但幾乎每次都差不多要等滿 15 分鐘, 難道每次到站牌, 都是公車剛好才開走嗎?

我們先看一所謂 "等待時間之詭論" (waiting time paradox)。假設公車依一參數爲 λ 之波松過程到站, 某人在時間 t 抵達此站。若以 γ_t 表至下一班公車來所需之等待時間 (即前面所提的剩餘壽命), 我們想求 γ_t 之期望值 $E(\gamma_t)$。則可能會有下述二矛盾的答案。

(a) 由指數分佈之無記憶性質, 導出等待時間之分佈應與自己到站牌的時間無關, 因此

$$E(\gamma_t)=E(\gamma_0)=\frac{1}{\lambda}。$$

(b) 由於某人抵達站牌之時間, 爲前後兩班車到站之區間中任意的一個點, 由對稱性知, 等待時間之期望值, 應大約是相鄰兩班車抵站牌之時間差距期望值之半, 即$1/(2\lambda)$。

這兩種論點看起來似乎皆很合理, 不過由之前所求出的波松過程之剩餘壽命的分佈, 我們知道 (a) 是對的。但 (b) 之錯誤何在?

原因爲雖然對一參數 λ 之波松過程, 每一到達間距 X_i 皆有相同之指數分佈, 且期望值爲 $1/\lambda$, 但若對一固定的 $t>0$, 找出 X_k 使得

$$\sum_{i=1}^{k-1}X_i<t\leq\sum_{i=1}^{k}X_i,$$

則此 X_k 之期望值, 並不等於 $1/\lambda$ 而是會大於 $1/\lambda$。且 $t\to\infty$ 時, 由 (14) 式知, 此期望值趨近至 $2/\lambda$, 而 $2/\lambda$ 之半即爲 $1/\lambda$。換句話說, 在 (b) 中我們誤將此一特別的 X_k (即總壽命) 之期望值也當做 $1/\lambda$。

我們可粗略地這樣想: 較長的區間比較短的區間有更大的機會包含一特定的點 t。

雖然間距皆有指數分佈, 且參數相同, 但由於是隨機變數, 區間就是有長有短, 假設在一條數線上有一些間距長短不一的點, 你隨機地取一點, 是不是較容易取自一間距較大的區間中? 同樣的道理, 若兩班公車抵站之間距很短, 你的到站時刻, 當然較不易落在其

中 (等車時間會較短), 而是較易落在一兩班車抵站間距較大的區間 (等車時間會較長)。在公車抵站間距爲指數分佈之假設下, 你等車時間之期望值 $1/\lambda$, 差不多等於在你之前與在你之後抵站之二車間距期望值

$$\frac{1}{\lambda}+\frac{1}{\lambda}(1-e^{-\lambda t})$$

之半, 而不是一般人以爲的 $1/\lambda$ 之半。

再看另一關於 "檢驗的詭論" (inspection paradox)。假設使用某電池, 一旦沒電了便立即換一同廠牌之新電池。電池壽命設爲 i.i.d., 這些換電池的時刻便構成一更新過程。

某品管人員想檢定電池之壽命, 原先與操作人員約好上午 8 時至工廠, 工廠在那時開始運作, 操作人員會在 8 時正啓用 30 個新電池。品管人員因故遲到, 但操作人員仍在 8 時正啓用電池。有些電池電耗盡了, 操作人員立即換新, 並記錄更換時間, 品管人員來廠後, 先登記正在測試的那些電池分別的啓用時間, 並等到每個電池電耗盡後, 拿到 30 個電池的壽命資料。結果取到的樣本其平均壽命差不多是原先以爲的電池壽命之兩倍, 這是怎麼回事?

假設電池之品質相同, 即壽命之分佈函數皆是某一同樣的 F, 我們以爲這些受測電池之壽命也同樣以 F 爲其分佈函數。其實並不然, 當 F 是指數分佈, 此例本質上與前面等公車的例子相同, 即此時正在受測電池的壽命分佈與 F 是不同的。這是值得重視的一個問題, 我們看到一個明顯偏差的檢驗計畫可能會導致錯誤的結論, 因爲我們所觀察到的樣本並非來自典型的母體 (typical population)。

在 Rao(1997) 一書中 pp.116-117, 亦提到一類似的例子。摩洛哥 (Morocco, 位於西北非洲之一國家) 之國立統計經濟應用研究所曾做一項研究, 目的是估計觀光客在他們國家之平均逗留時間。他們進行了兩種調查, 其一是對住在旅館的觀光客, 其二是對即將離境的旅客。從對 3,000 個在旅館的旅客之調查, 得平均逗留時間爲 17.8 日, 而從對 12,321 個離境的旅客之調查, 得平均逗留時間爲 9.0 日。前者差不多是後者的兩倍。你現在該知道何者才是可靠的停留時間之數據了。

7. 和其光同其塵

在老子第四章: 道沖, 而用之或不盈, 淵兮似萬物之宗。挫其銳, 解其紛, 和其光, 同其塵, 湛兮似或存, 吾不知誰之子, 象帝之先。

和其光同其塵的意思就是隱藏光耀, 混同塵俗。

不少統計學者均認爲波松分佈與常態分佈爲最重要的兩個分佈。但長久以來, 常態分佈具有獨尊的地位, 不要說波松分佈, 絕大部分的其他分佈, 其重要性往往被低估。

而在隨機過程裡, 波松過程與布朗運動 (Brownian Motion), 亦爲最重要且最基本的過程, 到處出現, 甚至在許多我們料不到的地方。但再度地, 波松過程的重要性也常被忽略。

忽視波松分佈或波松過程的重要性, 其實是統計學家的損失。因在很多離散的狀態

下, 波松分佈或波松過程常是最佳模式。波松分佈及波松過程, 彷彿和光同塵, 明珠卻蒙上灰塵。

雖然萬物有常, 但天下事物不能以常理度量的本就不少, 非常態分佈之處處存在, 也就不容懷疑。要發現波松分佈之重要, 遠比當伯樂容易的多, 只要你的眼光, 偶而離開常態分佈。

習題

1. 依第1節中關於豬肉換米的新聞報導裡, 所給米的重量, 估計 2^{64} 粒米的總重量, 並與全世界人口總重 (以每人平均 50 公斤重計) 相比。
2. 試估計將 F_{64} 這個數印出來, 約需多少本書?
3. 在第2節所述的致富方法中,
   (i) 若每月投資 $15,000$ 元, 40年後之本利和爲何?
   (ii) 若投資報酬率改爲每年 10%, 則 30 年後之本利和爲多少單位的錢? 40 年後呢?
4. 某工廠宣稱其生產的某型日光燈管可使用 $10,000$ 小時。某辦公室共安裝 32 支該型燈管, 使用後才一個月便壞了一支。該辦公室想了解這是否合理。假設燈管壽命爲 i.i.d. 之 $\mathcal{E}(\lambda)$ 分佈, 期望値爲 $10,000$ 小時, 又設每月上班 25 天, 每天開燈 8 小時。求
   (i) 某特定燈管1個月內會壞之機率;
   (ii) 辦公室1個月內至少會壞一支燈管之機率;
   (iii) 辦公室5個月內至少會壞二支燈管之機率。
5. 設 X 有 $\mathcal{E}(\lambda)$ 分佈, Y 爲一非負的隨機變數, 且與 X 獨立。試證 $P(X>Y+t \mid X>Y)=P(X>t)$。
6. 設 X 有自 1 開始, 參數 p 之幾何分佈, Y 爲一非負的隨機變數, 且與 X 獨立。試證 $P(X>Y+t \mid X>Y)=P(X>t)$。
7. 試證第 5 節中對一更新過程, 事件 $[A(t)=n]$ 與 $[S_n\le t<S_{n+1}]$ 等價, $\forall n\ge 0$。
8. 設 X 有 $\mathcal{E}(\lambda)$ 分佈, 又令 $[\cdot]$ 表最大整數函數。
   (i) 試證 $[X]$ 與 $X-[X]$ 獨立;
   (ii) 求 $X-[X]$ 之分佈。
9. 設 $\{A(t),\, t\ge 0\}$ 爲一參數爲 λ 之波松過程, 求總壽命 β_t 之分佈。

參考文獻

1. 黃文璋 (1995), 隨機過程, 華泰書局, 台北。
2. Cundy, H. M. (1966). Birds and atoms. *Math. Gazette* 50, 294-295.
3. Peterson, I. (1980). *Islands of Truth — A Mathematical Mystery Cruise*. W. H. Freeman and Company, New York.
4. Rao, C. R. (1997). *Statistics and Truth — Putting Chance to Work*, 2nd. ed. World Scientific Publishing Co. Pte. Ltd., Singapore.

—本文作者任教於國立高雄大學應用數學系—