

應用統計分佈於極值事件推估之教學文件

吳南靖

國立臺灣海洋大學 海洋環境資訊系

壹、前言

大自然的力量無遠弗屆，其極端表現，無論是劇烈降雨造成的淹水、海象劇變引發的港灣破壞，或是強陣風導致的設施損壞，都對人類社會構成嚴峻的挑戰。面對這些不可預測的極端事件，工程設計與風險管理絕不能掉以輕心。然而，如何在有限的歷史觀測資料下，推估未來可能出現的極端情境及其機率，正是應用數學與統計學所能發揮的價值。

在評估極端事件時，「重現期」(Return period) 是一個常用且關鍵的指標。它量化了某特定規模極端事件發生的稀有程度。它與年發生機率 P (即所謂的超越機率) 之間存在著簡潔的數學關係：

$$T=1/P \quad (1)$$

這意味著，在平穩條件下，若規模為 x_T 的事件(比如降雨強度、洪水流量、陣風風速、波高、...等)之重現期為 T 年，則任一年年最大值超過 x_T 之機率為 $1/T$ 。因此，理解並精確推估極端事件的發生機率及其對應的重現期，是確保工程安全與經濟效益的基石。

本文以真實氣象觀測資料為素材，設計一套可於課程實施之探究與實作流程(真實資料 → 步驟操作 → 作品產出與反思)，呼應當前教育界所倡議之素養導向與探究式學習。文章旨在成為連結學校教育與真實世界之教學資源，供數學、統計學教師及自學者參考，並回應學生「我們為什麼要學統計？」之核心提問。

貳、統計樣本之超越機率

在實務上，我們會從觀測資料或數值模擬結果中選取 N 個樣本。規模由大至小排序第 n 順位之值為 x_n 。利用 Goda (2000)歸納之排序公式，推估其累積機率

$$F_n = 1 - \frac{n-a}{N+b}, \quad n=1, 2, \dots, N \quad (2)$$

式中 a 和 b 分別為經驗參數。其選用值，常用的有 $a=0.5, b=0$ (Hazen, 1914), $a=0, b=1$ (Weibull, 1939), $a=0.375, b=0.25$ (Blom, 1958), 及 $a=0.44, b=0.12$ (Gringorten, 1963)。

關於取樣方法，一般以年最大法(annual maxima)從 K 年的資料中取每年之最大值，故 $N=K$ 。在統計樣本中，規模由大至小排序為 n 之事件的單次發生超越機率為

$$p=(1-F_n) \quad (3)$$

其年發生超越機率 P 就等於該事件單次發生之超越機率。即

$$P=p \quad (4)$$

逾閾值法 (Peaks-Over-Threshold, POT) 為相較於年最大法 (Annual Maxima, AM) 更具彈性之極值分析方法，能夠保留超過門檻值的所有極端事件資訊，不受一年只能取一值的限制。在某些特殊情形下，例如某區域常於一年內出現多次具災害性的洪水事件，若僅取年度最大值恐會忽略其他關鍵事件的風險貢獻，此時即適合改採逾閾值法來取樣。從 K 年的資料中篩選出 N 個事件時，平均每年有 λ 次事件 ($\lambda=N/K$)。則其年發生超越機率 P 為

$$P=1-(1-p)^\lambda \quad (5)$$

參、統計樣本與已知統計分佈之擬合

對工程設計而言，為確保結構物在使用年限內能發揮功能，實測資料的長度應滿足統計特性方能分析代表性的成果。然實際上往往因現有記錄不足而無法達成。此時，需要藉由統計方法，利用有限長度的資料，去推估不同重現期如 10 年、50 年、100 年等之設計條件。

在實際操作上，我們會計算累積機率 F_n 在某種已知的統計分佈所對應之推估值 \hat{x}_n ，並計算這些推估值與樣本值 x_n 之間的方均根差距(Root Mean Square Discrepancy)來評估樣本與統計分佈之擬合程度。

$$\text{RMSD}=\sqrt{\frac{1}{N}\sum_n^N(\hat{x}_n-x_n)^2} \quad (6)$$

RMSD 愈小表示擬合度愈高。從不同的統計分佈中，我們會挑選擬合度最高的，做為推估不同重現期設計條件之依據。另需注意， x_n 為樣本分位數而非真值，而 \hat{x}_n 則為分佈在同一累積機率下的理論分位數，因此式(6)衡量的是兩者之「擬合差距」(Discrepancy)，用以比較分佈擬合優劣，並非真值與推估結果之間的誤差(Error)。

肆、常用的統計分佈

在諸多既有的統計分佈中，本文僅介紹實務上常用 6 種的統計分佈：常態分佈(normal distribution)、對數常態分佈(log-normal distribution)、甘貝爾分佈(Gumbel distribution)、韋布爾分佈(Weibull distribution)、皮爾森第三型分佈(Pearson Type III distribution)、對數皮爾森第三型分佈(log-Pearson Type III distribution)。茲分述如下：

一、常態分佈(normal distribution)

使用常態分佈時，首先必須算出樣本之平均值 \bar{x} 與標準差 s 。

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (7)$$

$$s = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2} \quad (8)$$

要求出 F_n 所對應的推估值時，則先計算對應累積機率 F_n 之常態標準化值。

$$z_n = \Phi^{-1}(F_n) \quad (9)$$

其中 Φ^{-1} 為標準常態分布之反累積機率函數 (probit function)。然後依據以下公式求出推估值

$$\hat{x}_n = \bar{x} + z_n s \quad (10)$$

二、對數常態分佈(log-normal distribution)

使用對數常態分佈時，首先透過對數轉換，將所有的 x_n 轉換為 y_n 。

$$y_n = \log x_n \quad (11)$$

然後算出樣本經對數轉換後之平均值 \bar{y} 與標準差 s_y 。公式同(7)(8)，只不過變數換成 y_n 。

要針對 F_n 求出其所對應的推估值時，也是利用(9)計算其常態標準化值 z_n ，然後利用以下公式得 \hat{y}_n

$$\hat{y}_n = \bar{y} + z_n s_y \quad (12)$$

最後得到推估值

$$\hat{x}_n = 10^{\hat{y}_n} \quad (13)$$

三、甘貝爾分佈(Gumbel distribution)

又稱極端值第一型分佈(Extreme Value Type I distribution)。使用甘貝爾分佈時，一樣要利用式(7)及(8)計算平均值 \bar{x} 與標準差 s 。要針對 F_n 求出其所對應的推估值時，則先利用以下公式求出機率因子

$$k_n = \frac{\sqrt{6}}{\pi} (-\ln(-\ln(F_n)) - \gamma) \quad (14)$$

其中 $\gamma = 0.5772$ ，為歐拉·馬歇羅尼常數(Euler–Mascheroni constant)。然後推估值為

$$\hat{x}_n = \bar{x} + k_n s \quad (15)$$

四、韋布爾分佈(Weibull distribution)

又稱極端值第三型分佈(Extreme Value Type III distribution)。使用韋布爾分佈時，一樣要利用式(7)及(8)計算平均值 \bar{x} 與標準差 s 。不過，要求出 F_n 所對應的推估值時，則不需要先求出機率因子，而是直接由以下公式求出 \hat{x}_n

$$\hat{x}_n = \beta (-\ln(1 - F_n))^{\frac{1}{\alpha}} \quad (16)$$

其中 α 及 β 分別為韋布爾分佈的形狀參數及尺度參數，並無顯式公式，需以數值方法求解之。實務上，常利用以下公式(Justus et al., 1978)計算其近似值。

$$\alpha = (s / \bar{x})^{-1.086} \quad (17)$$

$$\beta = \bar{x} / \Gamma(1 + 1/\alpha) \quad (18)$$

其中， Γ 為 Gamma 函數。另 Rahman and Chattopadhyay (2020)建議把式(17)改為

$$\alpha = (0.9874 \bar{x} / s)^{1.0983} \quad (19)$$

可略為提升推估結果之準確性。

五、皮爾森第三型分佈(Pearson Type III distribution)

前四種分佈均為雙參數分佈，即只需用到樣本之平均值 \bar{x} (或 \bar{y}) 與標準差 s (或 s_y) 即可進行統計推估。皮爾森第三型分佈為三參數分佈，理論上具有更高之擬合度，但在樣本數極度缺乏的情形下，可能有過度描述的情形。除了平均值與標準差，皮爾森第三型分佈還需用到偏度。

$$c = \frac{1}{s^3 N} \sum_{n=1}^N (x_n - \bar{x})^3 \quad (20)$$

要求出 F_n 所對應的推估值時，皮爾森第三型分佈亦無顯式公式可用，需以數值方法求解之。實務上，最廣泛採用的方法是基於 Wilson and Hilferty (1931) 提出的轉換理論，它將皮爾森三型分佈的變數與常態標準化值 z_n 建立精確的近似關係。首先利用公式(9)求得 z_n ，接著透過以下由 Wilson and Hilferty (1931) 轉換所推導的頻率因子 k_n 公式求解：

$$k_n = \frac{2}{c} \left(\left(1 - \left(\frac{c}{6} \right)^2 + \left(\frac{c}{6} \right) z_n \right)^3 - 1 \right) \quad (21)$$

Kite (1977) 則將上式展開為多項式，其形式如下：

$$k_n = z_n + (z_n^2 - 1) \left(\frac{c}{6} \right) + \frac{1}{3} (z_n^3 - 6z_n) \left(\frac{c}{6} \right)^2 - (z_n^2 - 1) \left(\frac{c}{6} \right)^3 + z_n \left(\frac{c}{6} \right)^4 - \frac{1}{3} \left(\frac{c}{6} \right)^5 \quad (22)$$

最後將所求得之 k_n 代入式(15)以求出 \hat{x}_n 。

需留意式(21) 與式(22) 在數學上完全等價。早期為計算便利常採多項式展開之式(22)，現代教科書與軟體計算頻率因子時則多直接使用更為簡潔、能體現數學根源的閉合形式(21)。依據 Stedinger et al. (1993) 之說明，Wilson and Hilferty (1931) 及 Kite (1977) 公式僅適用於 $|z| \leq 2.33$ 且 $|c| \leq 2$ 的一般情況。若遇 $|z| > 2.33$ 或 $|c| > 2$ ，則應以數值法直接反解累積機率，所需之方程式與參數細節較繁，本文不再詳列，請參閱相關文獻(Wilson and Hilferty, 1931; Kite, 1977; Stedinger et al., 1993)。

六、對數皮爾森第三型分佈(log-Pearson Type III distribution)

對數皮爾森第三型分佈其實就是把所有的 x_n 做對數轉換得到 y_n ，然後把皮爾森第三型分佈公式裡的全部跟變數 x 有關的計算都換成對變數 y 做計算，做完 \hat{y}_n 之統計推估後再轉回 \hat{x}_n 之過程。它與把對數轉換套用到常態分佈變成對數常態分佈的過程如出一轍，故在此不展開贅述。

伍、設計條件之推估

在確定何種統計分佈與樣本資料的擬合度最佳之後，我們即可利用該統計分佈來進行設計條件之推估。如果我們的樣本資料是來自於年最大法取樣，要推估 T 年重現期之設計條件，就利用以下公式求其超越機率。

$$P=1/T \quad (23)$$

然後計算它在統計分佈裡的累積機率

$$F=1-P \quad (24)$$

最後利用前一節所述 6 種統計分佈由 F 求出理論值 \hat{x} 的步驟，則可得到該重現期之設計條件。

若採逾閾值法 (POT)，要推估 T 年重現期之設計條件，仍先由式(23)取 $P=1/T$ 作為年超越機率；但因一年平均有 λ 次事件，換算到分佈累積機率時改用式(25)：

$$F = (1 - P)^{\frac{1}{\lambda}} \quad (25)$$

求得累積機率 F 後，反推理論值 \hat{x} ，並視之為設計值。相關程序，亦與前一節所述之步驟完全相同。

陸、實例說明

我們蒐集中央氣象署自 1951 年至 2025 年於基隆氣象站所觀測到之年最大陣風風速 (詳表 1) 來做實例說明。假設我們搭著時光機回到 2020 年，故取 1951 年至 2020 年 (共 70 年) 之資料進行分析。首先，做資料排序，並做統計樣本累積機率之推估。各種公式所推估之統計樣本累積機率詳圖 1 所示。由圖可見，Hazen(1914)、Weibull(1939)、Blom(1958) 與 Gringorten(1963) 等繪圖機率 (plotting position) 公式，在本資料集的排序機率估計上所造成的差異相當有限；換言之，在相同排序與相同樣本數下，各公式所對應的累積機率

曲線幾乎重疊，對後續設計值推估的影響不大。基於「結果差異不顯著」以及「方法選擇需一致」兩項考量，本文後續計算遂採用年代較晚、且在極值分析情境中常用的 Gringorten (1963) 公式作為統一標準，以確保推估流程一致、便於比較與重現。

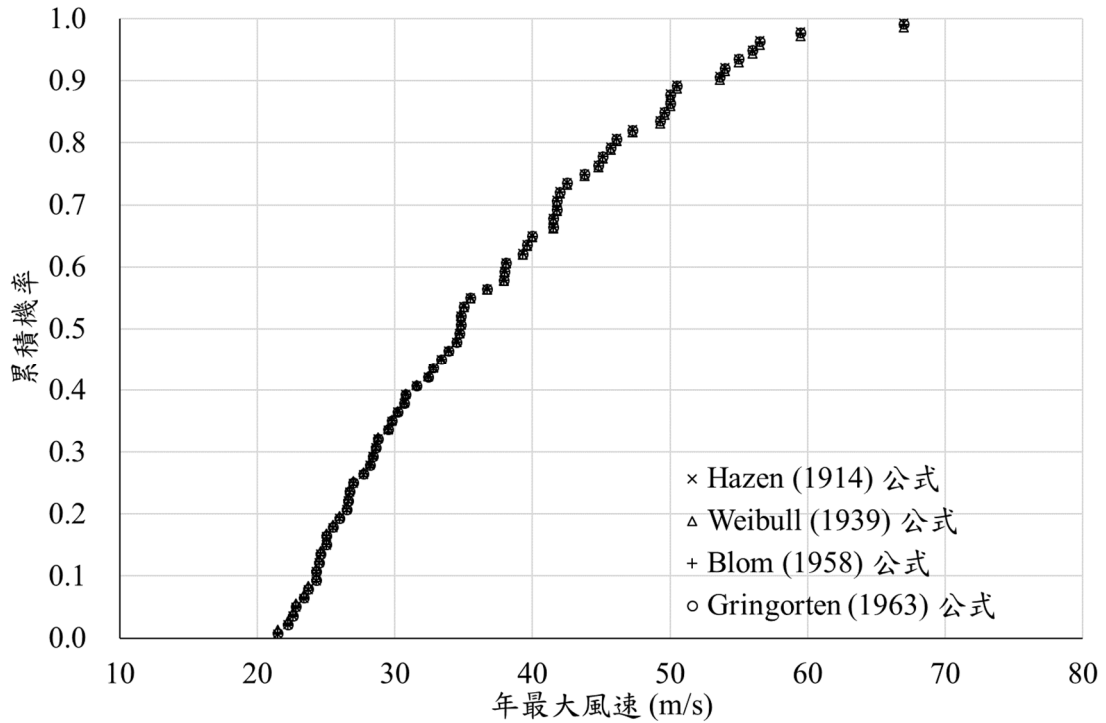


圖 1：年最大風速樣本於不同繪圖機率公式下之累積機率比較

表 1：中央氣象署基隆站歷年最大陣風風速

年份	年最大陣風風速	年份	年最大陣風風速	年份	年最大陣風風速
1951	30.2	1976	53.6	2001	44.8
1952	26.5	1977	56.5	2002	24.3
1953	42.0	1978	27.0	2003	33.4
1954	21.5	1979	26.6	2004	47.3
1955	25.0	1980	35.0	2005	49.3
1956	41.5	1981	33.9	2006	28.4
1957	30.8	1982	56.0	2007	49.6
1958	43.8	1983	24.5	2008	39.6
1959	54.0	1984	28.2	2009	31.6
1960	38.0	1985	46.1	2010	28.8
1961	42.5	1986	34.7	2011	22.6
1962	40.0	1987	41.8	2012	32.8
1963	45.1	1988	26.7	2013	50.0
1964	34.5	1989	25.0	2014	32.4
1965	55.0	1990	50.5	2015	45.7
1966	35.5	1991	27.7	2016	39.3
1967	30.7	1992	24.6	2017	41.5
1968	26.0	1993	23.4	2018	22.8
1969	50.0	1994	41.8	2019	28.6
1970	25.5	1995	23.7	2020	22.2
1971	67.0	1996	59.5	2021	23.9
1972	38.1	1997	34.8	2022	21.0
1973	24.3	1998	34.8	2023	24.8
1974	37.9	1999	29.5	2024	41.4
1975	29.8	2000	36.7	2025	27.3

接下來，我們算出排序第 1 至第 70 之 Gringorten (1963) 公式累積機率在前述的 6 種常用統計分佈之推估值 $\hat{x}_n (n=1, 2, \dots, 70)$ ，並計算它們與樣本值 x_n 之間的方均根差距，詳表 2 所示。統計樣本與各種分佈之累積機率比較則詳圖 2 所示。

表 2：各統計分佈之擬合差距指標 (RMSD) 比較

統計分佈	RMSD
常態分佈	2.48
對數常態分佈	1.47
甘貝爾分佈	1.65
韋布爾分佈	2.52
皮爾森第三型分佈	1.46
對數皮爾森第三型分佈	1.64

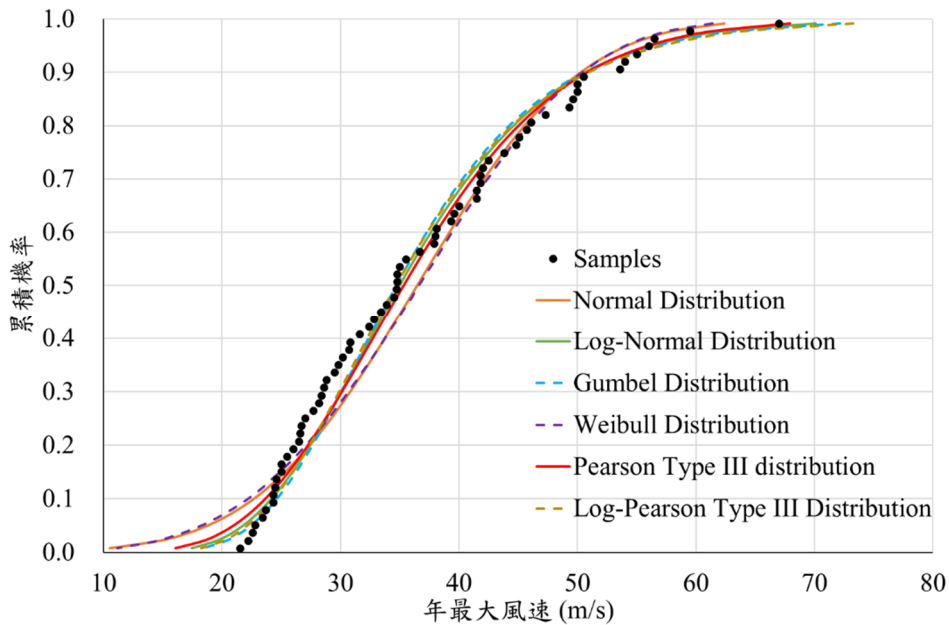


圖 2：統計樣本與各種分佈擬合結果之累積機率比較

經計算比對，與統計樣本擬合度最高的統計分佈為皮爾森第三型分佈。據此，我們計算 2 年、3 年、5 年、10 年等重現期之年最大風速分別為 35.3 m/s、40.1 m/s、45 m/s、50.8 m/s。由實際的 2021 至 2025 年資料，觀測到的最大風速為 41.4 m/s，故在這 5 年中，風速超過 35.3 m/s (2 年重現期)有發生，風速超過 40.1 m/s (3 年重現期)也有發生，風速超過 45 m/s (5 年重現期)及風速超過 50.8 m/s 則均未發生。

我們來分析結果之合理性。依據超越機率之定義，平均而言，未來每年年最大風速超過 35.3 m/s 的機率為 50%，超過 40.1 m/s 的機率為 33.33%，超過 45 m/s 的機率為 20%，超過 50.8 m/s 的機率為 10%。也就是說，拿 2020 年之前的分析結果來看 2021 至 2025 年：

會發生風速超過 35.3 m/s 的機率為 $1-(1-0.5)^5=96.9\%$ ；

會發生風速超過 40.1 m/s 的機率為 $1-(1-0.3333)^5=86.8\%$ ；

會發生風速超過 45 m/s 的機率為 $1-(1-0.2)^5=67.2\%$ ；

會發生風速超過 50.8 m/s 的機率為 $1-(1-0.1)^5=41\%$

整體而言，觀測結果與機率推估之趨勢一致：2 - 3 年重現期對應之門檻（35.3、40.1 m/s）在五年期間內出現的機率本就偏高，實際亦確有發生；至於較高門檻（45、50.8 m/s）雖然在五年內仍有一定機率可能出現，但其「未發生」的機率分別為 32.8%與 59%，因此 2021 - 2025 年未觀測到超越 45 m/s 與 50.8 m/s 並不矛盾。後續可隨資料年數增加再行檢驗其穩健性。因此，利用這套方法與流程所推得之 20 年、50 年等重現期之年最大陣風風速，應具相當高之合理性。

柒、結語

本文介紹了極端事件重現期的基本概念，說明如何透過統計樣本估算其超越機率，並闡述常用的 6 種統計分佈，包括其推估步驟與適用情境。實例部分以中央氣象署基隆氣象站自 1951 年至 2025 年之年最大陣風資料為例，進行統計分佈擬合，結果顯示皮爾森第三型分佈最為貼近歷史樣本。

需強調的是，本文所述為現行常見之極值統計分析方法，其目的在於介紹操作流程與應用方式。實際上，大氣與海象變化難以精準掌握，氣候變遷更進一步提高未來極端事件的不確定性。因此，本文所推算的各重現期對應之極端值僅供參考，並不代表未來必然會或不會發生，實務應用時仍應配合最新資料，結合領域專業知識進行整合判讀與評估。

參考文獻

- Blom, G. (1958). *Statistical estimates and transformed beta-variables*. John Wiley & Sons.
- Goda, Y. (2000). *Random seas and design of maritime structures (2nd ed.)*. World Scientific.
<https://doi.org/10.1142/3587>
- Gringorten, I. I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research*, 68(3), 813–814. <https://doi.org/10.1029/JZ068i003p00813>
- Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Transactions of the American Society of Civil Engineers*, 77, 1539–1640.
<https://doi.org/10.1061/taceat.0002563>
- Justus, C. G., Hargraves, W. R., Mikhail, A., & Graber, D. (1978). Methods for estimating wind speed frequency distributions. *Journal of Applied Meteorology*, 17(3), 350–353.
[https://doi.org/10.1175/1520-0450\(1978\)017<0350:MFEWSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1978)017<0350:MFEWSF>2.0.CO;2)
- Kite, G. W. (1977). *Frequency and risk analyses in hydrology*. Water Resources Publications.
- Rahman, S. M., & Chattopadhyay, H. (2020). A new approach to estimate the Weibull parameters for wind energy assessment: Case studies with four cities from the Northeast and East India. *International Transactions on Electrical Energy Systems*, 30(11), e12574.
<https://doi.org/10.1002/2050-7038.12574>
- Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In D. R. Maidment (Ed.), *Handbook of hydrology* (pp. 18.1–18.66). McGraw-Hill.
- Weibull, W. (1939). *A statistical theory of the strength of materials*. Generalstabens Litografiska Anstalts Förlag.
- Wilson, E. B., & Hilferty, M. M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America*, 17(12), 684–688.
<https://doi.org/10.1073/pnas.17.12.684>