詳細解答

STEP II 1991，第 16 題

概率與統計 • 二項分佈的正態近似

> **STEP II 1991，第 16 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/91-s2-q16

關鍵思路

追蹤水位差額（deficit）：即水位低於壩頂多少。排水令差額穩定上升，而每場驟雨則令它恰好下降 V；溢流就是差額變為負數。在兩天假期的問題中，只有一種降雨情況有可能引致溢流，而且屆時只有一個均勻分佈的時刻起作用。到了暑假，排水已經關掉，於是一切歸結為數算驟雨的次數，而這個次數服從二項分佈（binomial distribution）。

解答

第 (i) 部分：兩天假期

以日為單位，由假期開始的午夜起計算時間，並設 $D(t)$ 為時刻 t 的差額，故 $D(0) = V$。在兩場驟雨之間，D 以速率 u 上升，而每場驟雨令它下降 V。只要 D 一旦變為負數，水壩便會溢流。（我們按題目的原意假設，水庫存水充足，排水在這兩天內可以一直持續。）

設 R_1、R_2 分別為第 1 天和第 2 天下雨的事件，並設下雨的時刻（以一天的分數表示）為 T_1 和 T_2，兩者各自在 $[0,1]$ 上服從均勻分佈（uniform distribution）。各天互相獨立（independent），這一點已隱含在「任何一天下雨的概率為 p」這句話之中。

- **沒有下雨：**差額只會增加；不會溢流。
- **只有一天下雨**，時刻為 s：驟雨剛過後 $D = V + us - V = us \geq 0$，所以不會溢流（概率為零的事件 $s = 0$ 除外）。
- **兩天都下雨：**第一場驟雨過後 $D = uT_1 \geq 0$，這是安全的。第二場驟雨剛過後，即在時刻 $1 + T_2$，

$$D = V + u(1 + T_2) - 2V = u(1 + T_2) - V.$$

留意 T_1 已經消失：無論第一場驟雨何時落下，排水的速率都一樣。

所以，溢流恰好在兩天都下雨而且 $u(1+T_2) < V$ 時發生，即 $T_2 < \frac{V}{u} - 1$。若 $\frac{1}{2}V \le u \le V$，則 $\frac{V}{u} - 1$ 這個數位於 $[0,1]$ 之內，而由於 T_2 服從均勻分佈，

$$\mathrm{P}(\text{溢流}) = p^2\left(\frac{V}{u} - 1\right).$$

令上式等於 Q：

$$\frac{V}{u} = 1 + \frac{Q}{p^2} \quad\Longrightarrow\quad u = \frac{p^2 V}{p^2 + Q}.$$

條件 $Q < p^2$ 正是令這個結果前後一致的原因：由 $0 < Q/p^2 < 1$ 可得 $\frac{1}{2}V < u < V$，與假設相符。（沒有這個條件便無解。溢流的概率永遠不會超過 p^2，因為必須兩天都下雨。）

第 (i) 部分：暑假

現在 $u = 0$，所以差額永不增加。它由 kV 開始，每場驟雨令它下降 V，因此若 N 為 18 天之中下雨的日數，則

$$\text{溢流} \iff NV > kV \iff N > k.$$

各天互相獨立，每天下雨的概率為 $\frac{1}{3}$，所以 $N \sim \mathrm{B}(18, \frac{1}{3})$，其平均值為 6，方差（variance）為 $18 \cdot \frac{1}{3} \cdot \frac{2}{3} = 4$。

> **想一想** 這裏哪一種近似才算「合適」？為甚麼需要連續性修正（continuity correction）？

由於 $np = 6$ 而 $n(1-p) = 12$，正態近似（normal approximation）$N \approx \mathrm{N}(6,4)$ 是合理的。（泊松近似（Poisson approximation）則不合適，因為 p 並不小。）N 只取整數值，所以對整數 k 而言，事件 $N > k$ 就是事件 $N \ge k+1$，我們以 $X > k + \frac{1}{2}$ 來近似它，其中 $X \sim \mathrm{N}(6,4)$。我們需要

$$\mathrm{P}\left(Z > \frac{k + \frac{1}{2} - 6}{2}\right) < \frac{1}{10} \iff \frac{k - 5.5}{2} > 1.2816 \iff k > 8.06.$$

具體來說，$k = 8$ 給出 $\mathrm{P}(Z > 1.25) \approx 0.106$，太大了；而 $k = 9$ 給出 $\mathrm{P}(Z > 1.75) \approx 0.040$，符合要求。由於溢流與否只取決於 $N > k$ 是否成立，介乎 8 與 9 之間的非整數 k 並不比 $k = 9$ 好。所以工程師應把水庫排水至

$$k = 9,$$

即低於壩頂 $9V$。

第 (ii) 部分

現在共有 36 個半天，互相獨立，每個半天下雨的概率為 $\frac{1}{6}$。排水量仍然是零，所以溢流同樣表示 $N > k$，只是現在 $N \sim \mathrm{B}(36, \frac{1}{6})$。平均值仍然是 6，但方差是 $36 \cdot \frac{1}{6} \cdot \frac{5}{6} = 5$，比 4 大。

平均值相同時，方差較大即表示上尾較重：對每個高於平均值的 k，$\mathrm{P}(N > k)$ 都比之前大。所以所需的 k 較高。以 $\mathrm{N}(6, 5)$ 連同連續性修正作近似，我們需要

$$\frac{k - 5.5}{\sqrt{5}} > 1.2816 \iff k > 8.37,$$

而之前則是 $k > 8.06$。（這裏正態近似給出 $\mathrm{P}(N \geq 9) \approx 0.132$，而第 (i) 部分則是 0.106。）所以門檻上升了。若把 k 向上捨入至整數，兩種情況仍然都得出 $k = 9$。不過在 $k = 9$ 時，尾部概率現在約為 0.06 而不是 0.04，所以安全餘裕較小。

驗算

- 以蒙地卡羅模擬（Monte Carlo simulation）模擬兩天假期（400,000 次），取 $p = 0.5$、$Q = 0.1$ 及 $u = \frac{0.25}{0.35} V \approx 0.714V$，得出的溢流頻率為 0.0996。$p$ 和 Q 取其他數值時，結果同樣吻合。
- 極限情況：當 $Q \to p^2$ 時，$u \to \frac{1}{2}V$，這是兩天下雨仍可引致溢流的最慢排水速率。當 $Q \to 0$ 時，$u \to V$，在這個速率下，第二場驟雨永遠不會造成溢流。
- 精確的二項分佈尾部概率印證了正態近似。對 $\mathrm{B}(18, \frac{1}{3})$，$\mathrm{P}(N \geq 9) = 0.1076$ 及 $\mathrm{P}(N \geq 10) = 0.0433$。對 $\mathrm{B}(36, \frac{1}{6})$，$\mathrm{P}(N \geq 9) = 0.1332$ 及 $\mathrm{P}(N \geq 10) = 0.0652$。

常見錯誤

- 以為在兩天假期的問題中，第一場雨的時刻 T_1 有影響。其實它會互相抵消，因為排水速率是恆定的。
- 忘記連續性修正。沒有它，不等式便變成 $k > 8.56$，掩蓋了 N 是整數這個事實。
- 因為平均值相等，便說 (ii) 的答案「一樣」。其實控制尾部的是分佈的離散程度。

STEP 之外

雨水在隨機的時刻到來，而水則以穩定的速率排走，這是應用概率中的一個經典模型：儲存理論（storage theory），由 Moran 為研究水壩而發展出來。它與排隊（queues）和保險破產（insurance ruin）問題關係密切。在第 (ii) 部分，若把一天分割成愈來愈多、

愈來愈小的時段，同時保持平均值不變，則 $\mathrm{B}(n, 6/n)$ 趨向平均值為 6、方差為 6 的泊松分佈（Poisson distribution），而雨水的到來則成為一個泊松過程（Poisson process）。想進一步了解，可參閱 Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門）。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。