13.3 德國坦克問題

工作紙

題目

STEP III 1998，第 14 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/98-s3-q14

提示

請先自行嘗試解題。如果遇到困難，請逐一閱讀提示，每讀完一個便回到題目再試。標有停一停，試一試的提示是關鍵步驟。每個提示之後均留有空位供你演算。

想一想 設 M 是 $[0, N]$ 上 n 個獨立均勻分佈（uniform distribution）取值中的最大者，那麼 $\mathrm{P}(M \le m)$ 是多少？

階段 1 兩者皆為無偏

涵蓋：Z_1 和 Z_2 的平均值

想一想 Z_2 的平均值（mean）很快就能求出：每個 X_i 的平均值是多少？

停一停，試一試 至於 Z_1，先求最大值 M 的分佈（distribution）：何時有 $M \leq m$？求導得出其密度（density），再求 $\mathrm{E}(M)$。

階段 2　它們的方差

涵蓋：方差

想一想　先求 $\mathrm{E}(M^2)$，從而得出 $\mathrm{Var}(M)$，再按比例縮放得到 $\mathrm{Var}(Z_1)$。至於 Z_2，單獨一個 X_i 的方差（variance）是多少？各個 X_i 互相獨立（independent）。

階段 3 哪一個較好？

涵蓋：應該選用哪一個

想一想 把一個方差除以另一個。當 n 增大時，這個比值有何變化？這兩個估計量（estimator）之中，有沒有哪一個可能會小於你實際見過的某個序號？

背後的數學

充分性

為甚麼最大值好這麼多？一旦知道最大值 M，其餘各值不論 N 是多少，都只是在 $[0, M]$ 上均勻分佈：它們再沒有帶來關於 N 的任何信息。最大值是一個*充分統計量*（sufficient statistic）。C. R. Rao 與 David Blackwell 的一條定理（1945–47）指出，任何無偏估計量，只要換成它在給定充分統計量下的期望值，都可以得到改進，或至少不會變差；在這裏，這個做法正好把 Z_2 變成 Z_1。*Lehmann–Scheffé 定理*進一步證明，Z_1 是方差最小的無偏估計量。還要留意，Z_1 的方差以 $\frac{1}{n^2}$ 的速度下降，比大多數估計量的 $\frac{1}{n}$ 更快：這是因為 N 是分佈的端點，數據會擠向這個端點。

離散版本

真實的序號是整數 $1, 2, \ldots, N$，並且是以不放回（without replacement）方式抽樣的。這時最佳的無偏估計是 $\hat{N} = m + \frac{m}{k} - 1$，其中 m 是 k 個觀察到的序號中的最大者：即最大值加上各觀察值之間的平均間距。盟軍統計學家運用這類估計，主要根據變速箱和車輪的序號，估計德國的產量約為每月 246 輛坦克；相比之下，其後在德國紀錄中找到的數字是 245 輛，而情報報告所顯示的是 1400 輛。

歷史上的序號

此後這個方法曾多次派上用場，例如分析員曾根據最先售出的幾件產品的序號，估計某款新消費產品的生產數量。製造商汲取了教訓：現在不少製造商都會把序號編碼，

使人無法把它們當作簡單的順序編號來解讀。Richard Ruggles 與 Henry Brodie 於 1947 年在 *Journal of the American Statistical Association* 上記述了這項戰時工作。

延伸練習

1 **均方誤差（mean squared error）。**有偏估計量 M 滿足 $\mathrm{E}(M-N)^2 = \mathrm{Var}(M) + \big(\mathrm{E}(M)-N\big)^2$。計算此值，並與 $\mathrm{Var}(Z_1)$ 比較。求使 $\mathrm{E}(cM-N)^2$ 取最小值的常數 c。

2 **最小值也可以。**證明 $(n+1)\min\{X_i\}$ 也是 N 的無偏估計量，並求它的方差。這個估計量好用嗎？

3 **離散公式。**對於序號 $1,\ldots,N$，以不放回方式抽取 k 個，其中最大者為 m。證明 $\mathrm{P}(\max \le m) = \binom{m}{k}/\binom{N}{k}$，並驗證當 $k=1$ 時 $m+\frac{m}{k}-1$ 是無偏的。

4 **坦克。**若 5 輛繳獲的坦克的序號為 19、40、42、60、98，試估計坦克的生產數量。

延伸閱讀

- **Richard Ruggles & Henry Brodie, “An empirical approach to economic intelligence in World War II”**，*Journal of the American Statistical Association* 42 (1947)。最初的記述。
- **John Rice, *Mathematical Statistics and Data Analysis***（Duxbury），論及偏差、方差與充分性。
- **David Spiegelhalter, *The Art of Statistics***（Pelican）。

完整的引導式解答見第 13.3 節：
https://beyondstep.keithmcnulty.org/zh-hant/ch13/german-tanks.html

每星期一道 STEP 題目，完整講解：The STEP Up，以電郵免費發送，訂閱網址：keithmcnulty.substack.com