13.3 德国坦克问题

工作纸

题目

STEP III 1998，第 14 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/98-s3-q14

提示

请先自行尝试解题。如果遇到困难，请逐一阅读提示，每读完一个便回到题目再试。标有停一停，试一试的提示是关键步骤。每个提示之后均留有空位供你演算。

想一想 设 M 是 $[0, N]$ 上 n 个独立均匀分布（uniform distribution）取值中的最大者，那么 $\mathrm{P}(M \leq m)$ 是多少?

阶段 1 两者皆为无偏

涵盖：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：每个 X_i 的平均值是多少？

停一停，试一试 至于 Z_1，先求最大值 M 的分布（distribution）：何时有 $M \leq m$？求导得出其密度（density），再求 $\mathrm{E}(M)$。

阶段 2 它们的方差

涵盖：方差

想一想 先求 $\mathrm{E}(M^2)$，从而得出 $\mathrm{Var}(M)$，再按比例缩放得到 $\mathrm{Var}(Z_1)$。至于 Z_2，单独一个 X_i 的方差（variance）是多少？各个 X_i 互相独立（independent）。

阶段 3 哪一个较好?

涵盖：应该选用哪一个

> **想一想** 把一个方差除以另一个。当 n 增大时，这个比值有何变化？这两个估计量（estimator）之中，有没有哪一个可能会小于你实际见过的某个序号？

背后的数学

充分性

为甚么最大值好这么多？一旦知道最大值 M，其余各值不论 N 是多少，都只是在 $[0, M]$ 上均匀分布：它们再没有带来关于 N 的任何信息。最大值是一个*充分统计量*（sufficient statistic）。C. R. Rao 与 David Blackwell 的一条定理（1945–47）指出，任何无偏估计量，只要换成它在给定充分统计量下的期望值，都可以得到改进，或至少不会变差；在这里，这个做法正好把 Z_2 变成 Z_1。*Lehmann–Scheffé* 定理进一步证明，Z_1 是方差最小的无偏估计量。还要留意，Z_1 的方差以 $\frac{1}{n^2}$ 的速度下降，比大多数估计量的 $\frac{1}{n}$ 更快：这是因为 N 是分布的端点，数据会挤向这个端点。

离散版本

真实的序号是整数 $1, 2, \ldots, N$，并且是以不放回（without replacement）方式抽样的。这时最佳的无偏估计是 $\hat{N} = m + \frac{m}{k} - 1$，其中 m 是 k 个观察到的序号中的最大者：即最大值加上各观察值之间的平均间距。盟军统计学家运用这类估计，主要根据变速箱和车轮的序号，估计德国的产量约为每月 246 辆坦克；相比之下，其后在德国纪录中找到的数字是 245 辆，而情报报告所显示的是 1400 辆。

历史上的序号

此后这个方法曾多次派上用场，例如分析员曾根据最先售出的几件产品的序号，估计某款新消费产品的生产数量。制造商汲取了教训：现在不少制造商都会把序号编码，使人无法把它们当作简单的顺序编号来解读。Richard Ruggles 与 Henry Brodie 于 1947 年在 *Journal of the American Statistical Association* 上记述了这项战时工作。

延伸练习

1 **均方误差（mean squared error）。**有偏估计量 M 满足 $\mathrm{E}(M-N)^2 = \mathrm{Var}(M) + \big(\mathrm{E}(M) - N\big)^2$。计算此值，并与 $\mathrm{Var}(Z_1)$ 比较。求使 $\mathrm{E}(cM-N)^2$ 取最小值的常数 c。

2 **最小值也可以。**证明 $(n+1)\min\{X_i\}$ 也是 N 的无偏估计量，并求它的方差。这个估计量好用吗？

3 **离散公式。**对于序号 $1,\dots,N$，以不放回方式抽取 k 个，其中最大者为 m。证明 $\mathrm{P}(\max \le m) = \binom{m}{k}\big/\binom{N}{k}$，并验证当 $k=1$ 时 $m + \frac{m}{k} - 1$ 是无偏的。

4 **坦克。**若 5 辆缴获的坦克的序号为 19、40、42、60、98，试估计坦克的生产数量。

延伸阅读

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

完整的引导式解答见第 13.3 节：

https://beyondstep.keithmcnulty.org/zh-hans/ch13/german-tanks.html

每星期一道 STEP 题目，完整讲解：The STEP Up，以电子邮件免费发送，订阅网址：keithmcnulty.substack.com