详细解答

STEP I 1997，第 12 题

概率与统计 • 中心极限定理

STEP I 1997，第 12 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/97-s1-q12

关键思路

直接解出这个二次方程（quadratic equation），令 X 成为 T 的一个简单函数。这样，每个关于 X 的问题都变成关于 T 的问题，而后者的分布（distribution）我们完全清楚。一个巧妙的简化是：$1-T$ 同样在 $[0,1]$ 上服从均匀分布（uniform distribution）。最后一部分涉及大量独立（independent）副本之和，这正是提示我们使用中心极限定理（central limit theorem）。

解答

X 的分布

由二次公式（quadratic formula），$x^2+2x+T=0$ 的根为 $x=-1\pm\sqrt{1-T}$。由于 $0\le T\le 1$，这两个根都是实数，而较大的根取正号：

$$X=-1+\sqrt{1-T}.$$

当 T 取遍 $[0,1]$ 时，X 取遍 $[-1,0]$。

设 $U=1-T$。由于 T 在 $[0,1]$ 上服从均匀分布，U 亦然：把区间绕其中点反射，并不会改变均匀的密度（density）。所以 $X=\sqrt{U}-1$，其中 $U\sim \mathrm{U}[0,1]$。

$X>-\frac{1}{3}$ 的概率

由于 $X+1=\sqrt{U}$，而且两边皆为非负，

$$X>-\tfrac{1}{3}\iff \sqrt{U}>\tfrac{2}{3}\iff U>\tfrac{4}{9}.$$

因为两边皆为非负，平方后不等式的方向保持不变。因此

$$\mathrm{P}\left(X > -\tfrac{1}{3}\right) = \mathrm{P}\left(U > \tfrac{4}{9}\right) = 1 - \tfrac{4}{9} = \frac{5}{9}.$$

（等价地，$T < \frac{5}{9}$。）

平均值与方差

对于 $U \sim \mathrm{U}[0,1]$，其密度在 $[0,1]$ 上为 1，所以

$$\mathrm{E}(\sqrt{U}) = \int_0^1 u^{1/2}\,\mathrm{d}u = \frac{2}{3}, \qquad \mathrm{E}((\sqrt{U})^2) = \mathrm{E}(U) = \frac{1}{2}.$$

因此

$$\mathrm{E}(X) = \mathrm{E}(\sqrt{U}) - 1 = \frac{2}{3} - 1 = -\frac{1}{3}.$$

减去一个常数并不会改变方差（variance），所以

$$\mathrm{Var}(X) = \mathrm{Var}(\sqrt{U}) = \mathrm{E}(U) - (\mathrm{E}\sqrt{U})^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18},$$

正如所求。

> **想一想** 要求出这些矩（moments），我们需要 X 的概率密度函数（probability density function）吗？

不需要。对任意函数 g，$\mathrm{E}(g(U)) = \int_0^1 g(u)\,\mathrm{d}u$，所以我们可以直接用 U 来计算。如果你仍想知道密度：当 $-1 \le x \le 0$ 时，$\mathrm{P}(X \le x) = \mathrm{P}(U \le (x+1)^2) = (x+1)^2$，所以 X 在 $[-1,0]$ 上的密度为 $2(x+1)$。把 $x \cdot 2(x+1)$ 在该区间上积分，同样得到 $-\frac{1}{3}$。

800 个根之和

由于各次实验互相独立，各个 X_i 亦互相独立，而且每个的平均值（mean）为 $-\frac{1}{3}$，方差为 $\frac{1}{18}$。所以

$$\mathrm{E}(Y) = 800 \cdot \left(-\tfrac{1}{3}\right) = -\frac{800}{3}, \qquad \mathrm{Var}(Y) = 800 \cdot \frac{1}{18} = \frac{400}{9},$$

其中把方差相加时用到了独立性。Y 的标准差（standard deviation）为 $\frac{20}{3}$。

由于 800 很大，中心极限定理指出 Y 近似服从 $\mathrm{N}\left(-\frac{800}{3}, \frac{400}{9}\right)$。我们要求

$$\mathrm{P}(Y \le K) \approx \Phi\left(\frac{K + \frac{800}{3}}{20/3}\right) = 0.08.$$

查表得 $\Phi(1.405) \approx 0.92$，所以由对称性 $\Phi(-1.405) \approx 0.08$。因此

$$\frac{K + \frac{800}{3}}{20/3} \approx -1.405,$$

$$K \approx -\frac{800}{3} - 1.405 \times \frac{20}{3} \approx -266.67 - 9.37 \approx -276.$$

想一想 这里不是应该作连续性校正（continuity correction）吗?

不需要。每个 X_i 都是连续随机变量（continuous random variable），所以 Y 也是，因而无须作连续性校正。这种校正只用于以正态分布（normal distribution）近似离散分布（discrete distribution），例如二项分布（binomial distribution）。

验算

- 根是正确的：当 $T = 0$ 时，方程为 $x(x+2) = 0$，其较大的根为 $0 = -1 + \sqrt{1}$。
- $\mathrm{E}(X) = -\frac{1}{3}$ 位于 $[-1, 0]$ 之内，而密度 $2(x+1)$ 偏向 0 的一方，所以平均值高于中点 $-\frac{1}{2}$ 是合理的。
- 以 400 000 个 T 值进行模拟(simulation),得到 $\mathrm{P}(X > -\frac{1}{3}) \approx 0.555$,平均值 ≈ -0.3337,方差 ≈ 0.0556。直接模拟 Y 共 4000 次，得到 $\mathrm{P}(Y \leq -276.03) \approx 0.083$，与 0.08 吻合。

常见错误

- 在二次公式中取了负号。这样得到的是较小的根。
- 写成 $\mathrm{Var}(Y) = 800^2 \mathrm{Var}(X)$。这是 $800X_1$ 的方差，而不是 800 个独立副本之和的方差。
- 用了 $+1.405$ 而不是 -1.405。概率 0.08 位于下尾，所以 K 必定低于平均值。

STEP 之外

把 $X = g(U)$ 写成均匀变量 U 的函数，这个技巧也可以反过来用。如果 F 是任意一个连续的分布函数（distribution function），那么 $F^{-1}(U)$ 的分布函数就是 F。这个*逆变换法*（*inverse transform method*）正是电脑利用均匀随机数模拟随机变量的方法。中心极限定理在这里被当作黑箱使用，而在大学的第一个概率课程中，会用矩母函数（moment generating function）或特征函数（characteristic function）来证明它。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率入门教材）涵盖了这两个概念。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。