详细解答

STEP III 2007，第 14 题

概率与统计 · 几何概率

STEP III 2007，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/07-s3-q14

关键思路

包围所有洞孔的最小同心圆（或正方形），由最远的那个洞孔决定，所以我们要找的是最大值的分布，而 $P(\max \le r)$ 不过是单次投掷的概率（probability）的 n 次方。在圆盘上，一个洞孔与 O 的距离的平方 R^2 原来在 $[0,1]$ 上服从均匀分布（uniform distribution），于是一切都变得简洁。在正方形上，合适的“距离”是 $\max(|x|,|y|)$，它的表现与圆盘上的半径完全一样。这个巧合也令第 (iii) 部分几乎不费吹灰之力就解决了。

解答

第 (i) 部分

设 n 个洞孔与 O 的距离为 $R_1,\dots,R_n$。由于各次投掷互相独立，这些距离亦互相独立（independent）。对单次投掷及 $0 \le r \le 1$，洞孔与 O 的距离不超过 r，当且仅当它落在半径为 r 的圆盘内，所以

$$P(R_i \le r) = \frac{\pi r^2}{\pi \cdot 1^2} = r^2.$$

设 $U_i = R_i^2$。那么对 $0 \le u \le 1$，$P(U_i \le u) = P(R_i \le \sqrt{u}) = u$，所以每个 U_i 都在 $[0,1]$ 上服从均匀分布，而且各 U_i 互相独立。

以 O 为圆心并包围所有洞孔的最小圆，半径为 $M = \max_i R_i$，面积为 $\pi M^2 = \pi \max_i U_i$。记 $V = \max_i U_i$。由独立性，

$$P(V \le v) = P(U_1 \le v)\cdots P(U_n \le v) = v^n \qquad (0 \le v \le 1),$$

所以 V 的概率密度函数（density）为 nv^{n-1}，而

$$E(V) = \int_0^1 v \cdot nv^{n-1}\,\mathrm{d}v = \frac{n}{n+1}.$$

因此，包围全部 n 个洞孔的最小圆的期望（expected）面积为

$$\boxed{\frac{n\pi}{n+1}}.$$

至于包围最接近 O 的 $n-1$ 个洞孔的圆，其半径是第二大的距离，所以面积为 πW，其中 W 是 $U_1,\ldots,U_n$ 之中第二大的值。

> **想一想** 怎样求出第二大的值的分布?

$W \le w$ 当且仅当各 U_i 之中最多只有一个超过 w。要么没有任何一个超过（概率为 w^n），要么恰好有一个超过：选出是哪一个有 n 种方法，得出概率 $n(1-w)w^{n-1}$。所以

$$P(W \le w) = w^n + n(1-w)w^{n-1} = nw^{n-1} - (n-1)w^n.$$

我们不必求导，而可以利用 $E(W) = \int_0^1 P(W > w)\,\mathrm{d}w$，这条公式对任何取值于 $[0,1]$ 的随机变量（random variable）都成立：

$$E(W) = \int_0^1 (1 - nw^{n-1} + (n-1)w^n)\,\mathrm{d}w = 1 - 1 + \frac{n-1}{n+1} = \frac{n-1}{n+1}.$$

所以，包围最近的 $n-1$ 个洞孔的圆的期望面积为

$$\boxed{\frac{(n-1)\pi}{n+1}}.$$

（这符合一个漂亮的规律：n 个独立的均匀分布点把 $[0,1]$ 分成 $n+1$ 段间隙，每段的期望长度同为 $\frac{1}{n+1}$，所以第 k 小的点的平均值为 $\frac{k}{n+1}$。）

第 (ii) 部分

取通过 Q 且与镖靶各边平行的坐标轴，于是镖靶就是 $-1 \le x, y \le 1$。我们取包围洞孔的正方形，使其各边与镖靶的边平行。以 Q 为中心、半边长为 s 的正方形是 $\{|x| \le s,\ |y| \le s\}$，所以它包含洞孔 (x,y) 当且仅当 $\max(|x|,|y|) \le s$。

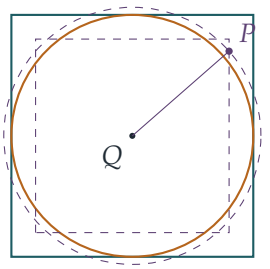
对第 i 个洞孔，设 $S_i = \max(|x_i|,|y_i|)$。对 $0 \le s \le 1$，

$$P(S_i \le s) = \frac{(2s)^2}{2^2} = s^2,$$

与第 (i) 部分的 $P(R_i \le r) = r^2$ 相同。包围所有洞孔的最小正方形，半边长为 $\max_i S_i$，面积为 $(2\max_i S_i)^2 = 4\max_i S_i^2$。与第 (i) 部分完全一样，各 S_i^2 互相独立，并在 $[0,1]$ 上服从均匀分布，所以它们的最大值的期望值为 $\frac{n}{n+1}$。包围所有洞孔的最小正方形的期望面积为

$$\boxed{\frac{4n}{n+1}}.$$

第 (iii) 部分



现在镖靶是正方形，而我们要找以 Q 为圆心并包围所有洞孔的最小圆。洞孔 $P=(x,y)$ 与 Q 的距离为 $\sqrt{x^2+y^2}$，而

$$\sqrt{x^2+y^2} \;\geq\; \max(|x|,|y|) = S,$$

等号仅在 P 位于坐标轴上时成立。（在图中，通过 P 的虚线圆永远不会小于通过 P 的虚线正方形的内切圆。）所以，若 D 是洞孔与 Q 的最大距离，则 $D \geq \max_i S_i$，而包围所有洞孔的圆的面积满足

$$\pi D^2 \;\geq\; \pi \max_i S_i^2.$$

由第 (ii) 部分，$\max_i S_i^2$ 的分布与圆形镖靶上 $\max_i R_i^2$ 的分布完全相同，所以

$$E(\pi D^2) \;\geq\; \pi\, E(\max_i S_i^2) = \frac{n\pi}{n+1},$$

这正是第 (i) 部分圆形镖靶的答案。这个不等式（inequality）是严格的：没有任何洞孔位于坐标轴上的概率为正，而这时 $D > \max_i S_i$。所以正方形镖靶上的圆，其期望面积**较大**。

直观地看：正方形镖靶包含了圆形镖靶（圆盘就是它的内切圆），而飞镖会散落到四个角落，所以最远的洞孔往往离中心更远。

验算

- 当 $n=1$ 时，公式给出 $\frac{\pi}{2}$ 和 2：均匀分布的 R^2 的平均值为 $\frac{1}{2}$，所以结果正确。当 $n\to\infty$ 时，面积趋向 π 和 4，即整个镖靶，这正合预期。
- 以蒙地卡罗模拟（Monte Carlo simulation）对 $n=2,3,5$ 各进行 2×10^5 次试验，结果与 $\frac{n\pi}{n+1}$、$\frac{(n-1)\pi}{n+1}$ 和 $\frac{4n}{n+1}$ 吻合至约三位有效数字；例如 $n=3$ 时得出 2.3561，而 $\frac{3\pi}{4}\approx 2.3562$。
- 对第 (iii) 部分，以数值积分（quadrature）计算精确分布，在 $n=3$ 时得出正方形镖靶上的期望面积约为 3.257，而圆盘上则为 2.356：一如所言，前者较大。

常见错误

- 先求 $E(M)$ 然后平方：期望面积是 $\pi E(M^2)$，而不是 $\pi\big(E(M)\big)^2$。
- 以为与中心的距离在 $[0,1]$ 上服从均匀分布。命中的概率与面积成正比，因此服从均匀分布的是 R^2 而不是 R；外围的环面积较大。
- 处理第二大的值时，忘记了因子 n，即哪一个洞孔落在圆外的选法数目。

STEP 之外

样本的最大值和第二大的值都是*顺序统计量*（order statistics）。对 $[0,1]$ 上 n 个独立的均匀分布变量，第 k 小的值服从贝塔分布（Beta distribution），平均值为 $\frac{k}{n+1}$。第 (iii) 部分是*随机优势*（stochastic dominance）（耦合，coupling）的一个例子：若一个随机变量总是不小于另一个，则它的期望值亦然，而这个比较无须任何细致的计算。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门）以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）都有讨论顺序统计量和随机变量的变换。

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。