详细解答

STEP I 1987，第 15 题

概率与统计 · 连续随机变量

STEP I 1987，第 15 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/87-s1-q15

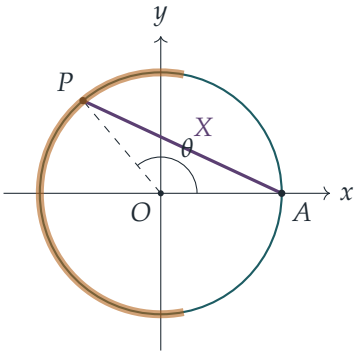
关键思路

“在圆上均匀分布（uniformly distributed）”的意思是 P 在圆心所成的角是均匀分布的，所以自然的做法是设 $P=(\cos\theta,\sin\theta)$，其中 θ 在 $[0,2\pi)$ 上均匀分布。这样，与 $(1,0)$ 的距离就是单位圆的一条弦（chord），而半角公式（half-angle formula）可把它化成单一个正弦。之后，平均值（mean）只需一个积分，方差（variance）几乎完全不用积分，而所求概率则变成一个关于角度区间的问题。

解答

设置

设 θ 为由正 x 轴到 OP 的角，其中 O 为原点。在圆上均匀地选取 P，即是均匀地选取弧长（arc length），所以 θ 在 $[0,2\pi)$ 上均匀分布，概率密度（density）为 $\frac{1}{2\pi}$。记 $A=(1,0)$ 及 $P=(\cos\theta,\sin\theta)$。



于是

$$X^2 = (\cos\theta - 1)^2 + \sin^2\theta = 2 - 2\cos\theta = 4\sin^2\tfrac{\theta}{2},$$

这里用了 $1 - \cos\theta = 2\sin^2\frac{\theta}{2}$。由于 $0 \le \frac{\theta}{2} < \pi$，我们有 $\sin\frac{\theta}{2} \ge 0$，所以取正平方根是稳妥的：

$$X = 2\sin\tfrac{\theta}{2}.$$

平均值

$$\mathrm{E}(X) = \frac{1}{2\pi}\int_0^{2\pi} 2\sin\tfrac{\theta}{2}\,\mathrm{d}\theta = \frac{1}{2\pi}\Big[-4\cos\tfrac{\theta}{2}\Big]_0^{2\pi} = \frac{1}{2\pi}(4+4) = \frac{4}{\pi}.$$

所以 $\mathrm{E}(X) = \dfrac{4}{\pi} \approx 1.273$。

方差

> **想一想** 要求 $\mathrm{E}(X^2)$，我们需要对 $4\sin^2\frac{\theta}{2}$ 积分吗？

不需要：未化成平方的形式更容易处理。$X^2 = 2 - 2\cos\theta$，而 $\cos\theta$ 在整整一圈上的平均值为零，因为 $\int_0^{2\pi}\cos\theta\,\mathrm{d}\theta = 0$。因此 $\mathrm{E}(X^2) = 2$，而

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 2 - \frac{16}{\pi^2} \approx 0.379.$$

X 大于其平均值的概率

我们需要 $2\sin\frac{\theta}{2} > \frac{4}{\pi}$，即 $\sin\frac{\theta}{2} > \frac{2}{\pi}$。留意 $0 < \frac{2}{\pi} < 1$，所以此不等式有解。设 $\alpha = \arcsin\frac{2}{\pi}$，它位于 $(0, \frac{\pi}{2})$ 内。当 $\frac{\theta}{2}$ 取遍 $[0,\pi)$ 时，正弦值由 0 升至 1 再回落，而它大于 $\sin\alpha$ 当且仅当

$$\alpha < \tfrac{\theta}{2} < \pi - \alpha, \qquad \text{即} \qquad 2\alpha < \theta < 2\pi - 2\alpha.$$

这就是图中加粗的圆弧：圆上位于 A 对面一侧的点。由于 θ 均匀分布，所求概率就是此区间的长度除以 2π：

$$\mathrm{P}(X > \mathrm{E}(X)) = \frac{2\pi - 4\alpha}{2\pi} = \boxed{1 - \frac{2}{\pi}\arcsin\frac{2}{\pi}} \approx 0.561.$$

验算

- X 的取值范围是 $[0,2]$，而平均值 $4/\pi \approx 1.27$ 位于中点 1 之上。这与图像吻合：在圆上颇大的一部分，弦长都接近其最大值（当 P 经过对面一侧时，X 变化缓慢），所以分布偏向较大的值。

- 基于同样原因，大于平均值的概率应略大于 $\frac{1}{2}$，而 0.561 正是如此。
- 在数值上，以中点法则（midpoint rule）对 2×10^5 个等距角度取平均，得到平均值 1.273240、方差 0.378861 及 $\mathrm{P}=0.56066$；而一个 10^6 点的模拟（simulation）在抽样误差范围内亦与此一致。

常见错误

- 取 $\theta\in(-\pi,\pi]$ 而仍写 $X=2\sin\frac{\theta}{2}$。此时 $\sin\frac{\theta}{2}$ 可以是负数，你需要写 $X=2\left|\sin\frac{\theta}{2}\right|$；使用 $[0,2\pi)$ 就可避免绝对值（modulus）。
- 把 P 的 x 坐标当作在 $[-1,1]$ 上均匀分布。在圆上均匀分布是指在角度（弧长）上均匀，而不是在 x 上均匀。

STEP 之外

这是随机变量变换（transformation of a random variable）的第一个例子：X 的概率密度函数是 $[0,2]$ 上的 $f(x)=\dfrac{2}{\pi\sqrt{4-x^2}}$，属于反正弦（arcsine）类型的分布，而上面每一个答案都可以由它重新推导出来。这道题亦暗示了“随机”一词可以多么难以捉摸：以不同的自然方式随机选取圆的一条弦，会得出不同的答案（贝特朗悖论（Bertrand's paradox））。这两个概念在 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）中都有进一步发展。

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。