详细解答

STEP II 1999，第 13 题

概率与统计 · 连续随机变量

> **STEP II 1999，第 13 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/99-s2-q13

关键思路

这道题虽然归入力学（mechanics）一类，其实是一道关于连续随机变量（continuous random variable）的题目。只要留意到较短的一段有简单的分布，一切便迎刃而解：若折断点在棒上均匀分布（uniform），则较短的一段在棒长的 $[0, \frac{1}{2}]$ 上均匀分布。比值 R 是较短一段的递增函数，所以“$R < r$”可化为只涉及较短一段的命题。由此得出分布函数（distribution function），再求导便得到密度（density）。

解答

R 的分布

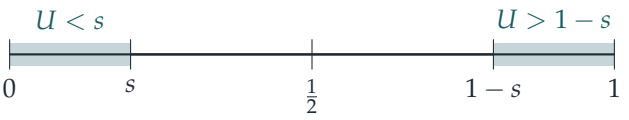
设棒长为 1；比值与所用单位无关。设折断点与其中一端的距离为 U，其中 U 在 $[0,1]$ 上均匀分布。两段的长度分别为 U 和 $1-U$，所以较短一段的长度为

$$S = \min(U, 1-U),$$

而较长一段的长度为 $1-S$。于是 $0 \le S \le \frac{1}{2}$，并且

$$R = \frac{S}{1-S}.$$

对于 $0 \le s \le \frac{1}{2}$，事件 $S < s$ 恰好在 $U < s$ 或 $U > 1-s$ 时发生。这是两个互不相交（disjoint）的区间，总长度为 $2s$，所以 $\mathrm{P}(S < s) = 2s$。换句话说，S 在 $[0, \frac{1}{2}]$ 上均匀分布。



现设 $0 \le r \le 1$。函数 $s \mapsto \dfrac{s}{1-s}$ 在 $[0, \frac{1}{2}]$ 上严格递增（strictly increasing）（其导数为 $1/(1-s)^2 > 0$），所以我们可以把不等式反解出来：

$$R < r \iff \frac{S}{1-S} < r \iff S < r(1-S) \iff S < \frac{r}{1+r},$$

其中两边同乘 $1-S$ 是稳妥的，因为 $1-S \ge \frac{1}{2} > 0$。当 $0 \le r \le 1$ 时，$\dfrac{r}{1+r}$ 的值落在 $[0, \frac{1}{2}]$ 之内，所以

$$\boxed{\mathrm{P}(R < r) = \frac{2r}{1+r} \qquad (0 \le r \le 1).}$$

由于较短的一段永远不会比较长的一段长，R 必定落在 $[0,1]$ 之内。所以当 $r \le 0$ 时 $\mathrm{P}(R < r) = 0$，当 $r \ge 1$ 时 $\mathrm{P}(R < r) = 1$，这与公式在 $r = 1$ 处取值 1 吻合。

? **想一想** 为甚么不直接用 U 来做，即取 $R = U/(1-U)$?

这只在 $U \le \frac{1}{2}$ 时才正确。当 $U > \frac{1}{2}$ 时，较短的一段是 $1-U$，而 $R = (1-U)/U$。你可以分成这两种情况来处理，根据对称性，每种情况各贡献 $\frac{r}{1+r}$。改用 S 则一次过把这些细节处理妥当。

密度、平均值与方差

在 $(0,1)$ 上对分布函数 $F(r) = \dfrac{2r}{1+r} = 2 - \dfrac{2}{1+r}$ 求导，得到概率密度函数（probability density function）

$$f(r) = \frac{2}{(1+r)^2} \quad (0 \le r \le 1), \qquad f(r) = 0 \text{ 其他情况.}$$

验算一下：$\int_0^1 f(r)\,\mathrm{d}r = F(1) - F(0) = 1$。

求平均值（mean）时，写成 $r = (1+r) - 1$：

$$\begin{aligned}
\mathrm{E}(R) &= \int_0^1 \frac{2r}{(1+r)^2}\,\mathrm{d}r = 2\int_0^1 \left(\frac{1}{1+r} - \frac{1}{(1+r)^2}\right)\mathrm{d}r \\
&= 2\Big[\ln(1+r) + \frac{1}{1+r}\Big]_0^1 = 2\left(\ln 2 + \tfrac{1}{2} - 1\right) \\
&= 2\ln 2 - 1 \approx 0.386.
\end{aligned}$$

求二阶矩（second moment）时，写成 $r^2 = (1+r)^2 - 2(1+r) + 1$，于是

$$\frac{r^2}{(1+r)^2} = 1 - \frac{2}{1+r} + \frac{1}{(1+r)^2}.$$

那么

$$\begin{aligned}\mathrm{E}(R^2) &= 2\int_0^1 \left(1 - \frac{2}{1+r} + \frac{1}{(1+r)^2}\right) \mathrm{d}r = 2\left[r - 2\ln(1+r) - \frac{1}{1+r}\right]_0^1 \\ &= 2\left(1 - 2\ln 2 - \tfrac{1}{2} + 1\right) = 3 - 4\ln 2.\end{aligned}$$

因此方差（variance）为

$$\begin{aligned}\mathrm{Var}(R) &= \mathrm{E}(R^2) - (\mathrm{E}(R))^2 = 3 - 4\ln 2 - (4(\ln 2)^2 - 4\ln 2 + 1) \\ &= 2 - 4(\ln 2)^2 \approx 0.0782.\end{aligned}$$

含 $\ln 2$ 的项互相抵消，剩下这个简洁的形式。

验算

- 在 $r = 1$ 处概率为 1，理应如此。在 $r = \frac{1}{2}$ 处概率为 $\frac{2}{3}$：较长一段超过较短一段的两倍，恰好是折断点落在棒两端各三分之一的范围内，而这两部分的总长度为 $\frac{2}{3}$。
- 平均值 0.386 介乎 0 与 1 之间，并且小于 $\frac{1}{2}$，这与密度由 $r = 0$ 处的 2 递减至 $r = 1$ 处的 $\frac{1}{2}$ 相符。
- 模拟 10^6 次随机折断，得到样本平均值 0.3862 及样本方差 0.0783，相比之下理论值为 $2\ln 2 - 1 = 0.38629$ 及 $2 - 4(\ln 2)^2 = 0.07819$。

常见错误

- 对所有 U 都取 $R = U/(1-U)$。这个量可以大于 1，所以它并不是较短一段与较长一段之比。
- 忘记折断点可以靠近任何一端，因而得出 $\mathrm{P}(R < r) = \frac{r}{1+r}$。
- 用较短一段的平均值 $\frac{1}{4}$ 来得出 $\mathrm{E}(R) = \frac{1/4}{3/4} = \frac{1}{3}$。一般而言 $\mathrm{E}\big(g(S)\big) \neq g\big(\mathrm{E}(S)\big)$。

STEP 之外

这里用的是变换随机变量的标准方法，即“分布函数法”。若 $Y = g(X)$ 而 g 为递增函数，则 $F_Y(y) = F_X\big(g^{-1}(y)\big)$，求导后便得到变量代换公式（change-of-variables formula）$f_Y(y) = f_X\big(g^{-1}(y)\big)\left|\mathrm{d}g^{-1}/\mathrm{d}y\right|$。同一想法可以引申到顺序统计量（order statistics），以及经典的折棒问题，例如三段能构成三角形的概率。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门）是研读连续随机变量变换的好去处。

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。