详细解答

STEP II 2006，第 14 题

概率与统计 · 连续随机变量

STEP II 2006，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/06-s2-q14

关键思路

整道题都建基于一个原函数 (antiderivative)。分母 $x\ln x$ 是 $\ln x$ 乘以 x，而 $1/x$ 正是 $\ln x$ 的导数，所以被积函数具有 $g'(x)/g(x)$ 的形式，其中 $g(x)=\ln x$。因此

$$\int \frac{1}{x\ln x}\,\mathrm{d}x = \ln|\ln x| + C.$$

此后，每一部分都只需运用“总概率为 1”，并小心留意正负号，以及密度（density）实际上在哪里不为零。

解答

草图

设 $y=\dfrac{1}{x\ln x}$，其中 $x>0$，$x\neq 1$。

正负号。由于 $x>0$，y 的正负号与 $\ln x$ 相同：在 $0<x<1$ 上 $y<0$，在 $x>1$ 上 $y>0$。曲线与两条坐标轴都不相交。

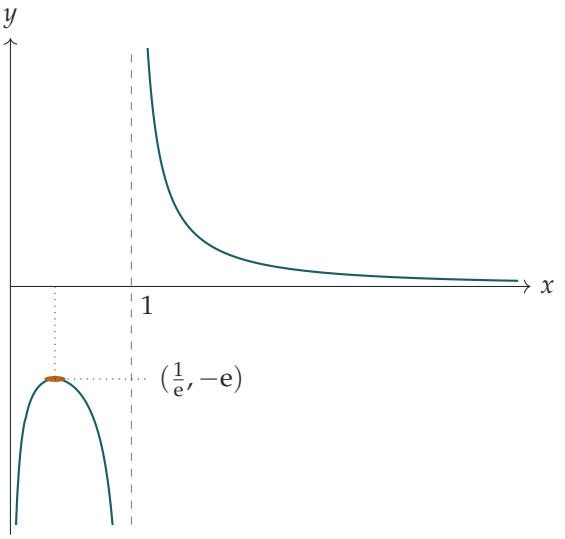
两端的变化。

- 当 $x\to 0^+$ 时：题目告诉我们 $x\ln x\to 0$，而且在该处 $x\ln x<0$，所以 $y\to -\infty$。
- 当 $x\to 1$ 时：$x\ln x\to 0$，从左边趋近时为负，从右边趋近时为正，所以当 $x\to 1^-$ 时 $y\to -\infty$，当 $x\to 1^+$ 时 $y\to +\infty$。直线 $x=1$ 是垂直渐近线（vertical asymptote）。
- 当 $x\to\infty$ 时：$x\ln x\to\infty$，所以 $y\to 0^+$，而 x 轴是渐近线。

转向点（*turning points*）。由商法则（quotient rule）或链式法则（chain rule），

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\ln x + 1}{(x \ln x)^2},$$

它只在 $\ln x = -1$ 时为零，即 $x = 1/\mathrm{e}$，此时 $y = \dfrac{1}{(1/\mathrm{e})(-1)} = -\mathrm{e}$。当 $0 < x < 1/\mathrm{e}$ 时，我们有 $\ln x + 1 < 0$，所以 y 递增；当 $x > 1/\mathrm{e}$（且 $x \neq 1$）时，函数递减。因此 $(1/\mathrm{e}, -\mathrm{e})$ 是左边分支的局部极大值（local maximum）点，而右边分支则由 $+\infty$ 一直递减至 0。



第 (i) 部分

这里 $a = \frac{1}{4}$ 而 $b = \frac{1}{2}$，两者都在 $(0,1)$ 内，而在该区间上 $x \ln x < 0$。利用上面的原函数，

$$\begin{aligned}\int_{1/4}^{1/2} \frac{\mathrm{d}x}{x \ln x} &= \Big[\ln|\ln x|\Big]_{1/4}^{1/2} = \ln|\ln \tfrac{1}{2}| - \ln|\ln \tfrac{1}{4}| \\ &= \ln(\ln 2) - \ln(2 \ln 2) = -\ln 2.\end{aligned}$$

总概率必须为 1，所以 $\lambda \cdot (-\ln 2) = 1$，由此得

$$\lambda = -\frac{1}{\ln 2}.$$

这个负号正是我们所需要的：在 $[\frac{1}{4}, \frac{1}{2}]$ 上 $x \ln x < 0$，所以 λ 必须为负，f 在该处才会非负。

第 (ii) 部分

当 $\lambda = 1$ 时, 密度必须非负, 所以区间 $[a, b]$ 必须位于 $x \ln x > 0$ 的范围内; 由于 $b > a > 1$, 这一点确实成立。于是 $\ln a, \ln b > 0$, 并且

$$1 = \int_a^b \frac{\mathrm{d}x}{x \ln x} = \ln(\ln b) - \ln(\ln a) = \ln \frac{\ln b}{\ln a}.$$

所以 $\dfrac{\ln b}{\ln a} = \mathrm{e}$, 即 $\ln b = \mathrm{e} \ln a = \ln\left(a^{\mathrm{e}}\right)$; 又因为 ln 是一对一 (one-to-one) 的, 故 $b = a^{\mathrm{e}}$, 正如所求。

第 (iii) 部分

取 $a = \mathrm{e}$, 由第 (ii) 部分得 $b = \mathrm{e}^{\mathrm{e}}$。由于 $1 < \frac{3}{2} < 2 < \mathrm{e}$, 整个区间 $[\mathrm{e}^{3/2}, \mathrm{e}^2]$ 都位于 $[a, b] = [\mathrm{e}, \mathrm{e}^{\mathrm{e}}]$ 之内, 所以

$$\mathrm{P}(\mathrm{e}^{3/2} \le X \le \mathrm{e}^2) = \Big[\ln(\ln x)\Big]_{\mathrm{e}^{3/2}}^{\mathrm{e}^2} = \ln 2 - \ln \tfrac{3}{2} = \ln \tfrac{4}{3}.$$

要把这个结果化成分数, 可使用麦克劳林级数 (Maclaurin series) $\ln(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} - \frac{t^4}{4} + \cdots$, 取 $t = \frac{1}{3}$, 并保留四项:

$$\ln \tfrac{4}{3} \approx \frac{1}{3} - \frac{1}{18} + \frac{1}{81} - \frac{1}{324} = \frac{108 - 18 + 4 - 1}{324} = \frac{93}{324} = \frac{31}{108}.$$

? **想一想** 我们怎样知道取到第四项就停下来是合理的?

这个级数正负项交替出现, 而且各项的大小递减, 所以取四项之后的误差小于第五项, 即 $\frac{1}{5 \cdot 3^5} = \frac{1}{1215} < 0.001$。因此 $\mathrm{P} = \ln \frac{4}{3} \approx \frac{31}{108}$, 正如所求。

第 (iv) 部分

现在 $a = \mathrm{e}^{1/2}$, 所以由第 (ii) 部分得

$$b = a^{\mathrm{e}} = \mathrm{e}^{\mathrm{e}/2}.$$

? **想一想** 在积分之前, 先想想: 密度实际上在哪里不为零?

由于 $\mathrm{e} < 3$, 我们有 $\mathrm{e}/2 < \frac{3}{2}$, 所以 $b = \mathrm{e}^{\mathrm{e}/2} < \mathrm{e}^{3/2}$。区间 $[\mathrm{e}^{3/2}, \mathrm{e}^2]$ 完全位于 b 的右边, 而在那里 $\mathrm{f}(x) = 0$。因此

$$\mathrm{P}(\mathrm{e}^{3/2} \le X \le \mathrm{e}^2) = 0.$$

(若不加思索地把 $1/(x \ln x)$ 由 $\mathrm{e}^{3/2}$ 积分至 e^2, 便会再次得到 $\ln \frac{4}{3}$, 而这是错的。)

验算

- 以数值计算，$\mathrm{e}^{-1} \approx 0.368$，而在 $(0,1)$ 上逐点检视 y 的值，可见它在该区间上的最大值 $\approx -2.718 = -\mathrm{e}$。
- 辛普森法则（Simpson's rule）给出 $\int_{1/4}^{1/2} \frac{\mathrm{d}x}{x \ln x} \approx -0.6931 = -\ln 2$，所以 $\lambda \approx -1.4427$。
- 对于若干个 $a > 1$ 的值（例如 $a = 2$、$a = 7.3$），$\int_a^{a^{\mathrm{e}}} \frac{\mathrm{d}x}{x \ln x}$ 的结果都是 1.0000。
- $\ln \frac{4}{3} \approx 0.28768$ 而 $\frac{31}{108} \approx 0.28704$，两者相差不超过 0.0007。
- 在 (iv) 中，$b = \mathrm{e}^{\mathrm{e}/2} \approx 3.89$，而 $\mathrm{e}^{3/2} \approx 4.48$。

常见错误

- 在第 (i) 部分写成 $\ln(\ln x)$，但那里 $\ln x < 0$：原函数需要加上绝对值（modulus），即 $\ln|\ln x|$。
- 答 $\lambda = \frac{1}{\ln 2}$：正的 λ 会令“密度”变成负数。
- 在第 (iv) 部分，没有先检查 $[\mathrm{e}^{3/2}, \mathrm{e}^2]$ 是否位于支撑集（support）$[a,b]$ 之内，便直接在这个区间上积分。

STEP 之外

代换（substitution）$u = \ln x$ 把这个密度变成 $[\ln a, \ln b]$ 上的 $1/u$，而条件 $b = a^{\mathrm{e}}$ 其实就是 $\int_{\ln a}^{\ln b} \frac{\mathrm{d}u}{u} = 1$。像这样对随机变量（random variable）作变换（这里是 $U = \ln X$），是大学一年级概率论的核心工具；在那里，随机变量的函数的密度是由链式法则求出的。同一个函数 $1/(x \ln x)$ 在分析学中也十分有名：$\sum 1/(n \ln n)$ 发散（diverges），这可以通过与这个积分比较而得出，尽管它的各项趋向零。关于随机变量的变换，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*。

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。