详细解答

STEP II 1988，第 16 题

概率与统计 · 正态分布

STEP II 1988，第 16 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/88-s2-q16

关键思路

关于两根的一切，都由单一个随机变量（random variable）B 决定，所以每一个关于 X_1 和 X_2 的问题，都应该转化为“B 落在哪个区间”的问题。判别式（discriminant）告诉我们两根何时为实数，而韦达定理（Vieta's formulas）（两根之和为 $-2B$，两根之积为 1）令 (i) 和 (ii) 的条件很容易转换。之后，每个答案都是标准正态变量（standard normal variable）的条件概率（conditional probability）或条件期望值（conditional expectation），而求期望值只需计算一个简短的积分。

解答

全题设 $B \sim \mathrm{N}(0,1)$，其密度函数（density）为 $\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}$，分布函数（distribution function）为 Φ。设 R 为“两根皆为实数”这个事件。

实根

$X^2 + 2BX + 1$ 的判别式为 $4B^2 - 4$，所以两根为实数当且仅当 $B^2 \geq 1$，即 $|B| \geq 1$。由正态分布（normal distribution）的对称性，

$$\mathrm{P}(R) = \mathrm{P}(|B| \geq 1) = 2(1 - \Phi(1)) = 2(1 - 0.84134) = 0.31731\ldots \approx 0.317.$$

（是否包括等根的情况 $|B| = 1$ 并无分别，因为它的概率为零。）

当两根为实数时，它们是

$$X_{1,2} = -B \pm \sqrt{B^2 - 1}, \qquad X_1 + X_2 = -2B, \qquad X_1 X_2 = 1.$$

第 (i) 部分

想一想 怎样把“两根都大于 $\frac{1}{5}$”转化为关于 B 的条件?

首先，两根必须都是正数。两根之积为 $1>0$，所以它们同号，而这个符号就是两根之和 $-2B$ 的符号。因此我们需要 $B<0$，连同 R 即表示 $B\le -1$。记 $b=-B\ge 1$；这时两根为 $b\pm\sqrt{b^2-1}$，都是正数。

由于两根之积为 1，较小的根是较大的根的倒数：

$$b-\sqrt{b^2-1}=\frac{1}{b+\sqrt{b^2-1}}.$$

所以较小的根大于 $\frac{1}{5}$ 当且仅当较大的根小于 5，而这时较大的根自然也大于 $\frac{1}{5}$。因此条件就是

$$\sqrt{b^2-1}<5-b.$$

这需要 $b<5$，而当 $1\le b<5$ 时两边皆为非负数，所以可以把两边平方而不改变不等式：

$$b^2-1<25-10b+b^2 \iff 10b<26 \iff b<2.6.$$

（$b\ge 5$ 的值无论如何都不符合，因为 $2.6<5$。）因此，在两根为实数的条件下，两根都大于 $\frac{1}{5}$ 当且仅当 $-2.6<B\le -1$，于是

$$\begin{aligned}\mathrm{P}(X_1,X_2>\tfrac{1}{5}\,|\,R)&=\frac{\mathrm{P}(-2.6<B\le -1)}{\mathrm{P}(|B|\ge 1)}=\frac{\Phi(2.6)-\Phi(1)}{2(1-\Phi(1))}\\&=\frac{0.99534-0.84134}{0.31731}=0.4853\ldots\approx 0.485.\end{aligned}$$

第 (ii) 部分

由于 $|X_1+X_2|=2|B|$，我们要求的是 $\mathrm{E}(2|B|\,||B|\ge 1)$。要计算给定某事件下的条件期望值，我们在该事件上积分，再除以该事件的概率：

$$\mathrm{E}(|B|\,||B|\ge 1)=\frac{1}{\mathrm{P}(|B|\ge 1)}\int_{|z|\ge 1}|z|\,\phi(z)\,\mathrm{d}z=\frac{2\int_1^\infty z\,\phi(z)\,\mathrm{d}z}{2(1-\Phi(1))},$$

这里用到了 ϕ 的对称性。这个积分很容易计算，因为 $z\,\phi(z)$ 正是 $\phi(z)$ 的导数的负值：

$$\int_1^\infty z\,\phi(z)\,\mathrm{d}z=\Big[-\phi(z)\Big]_1^\infty=\phi(1)=\frac{\mathrm{e}^{-1/2}}{\sqrt{2\pi}}=0.24197\ldots$$

因此

$$\mathrm{E}(|X_1+X_2|\,|\,R)=\frac{2\phi(1)}{1-\Phi(1)}=\frac{2\times 0.24197}{0.15866}=3.050\ldots\approx 3.05.$$

验算

- 试取 $B = -2.6$：方程为 $X^2 - 5.2X + 1 = 0$，两根恰好是 5 和 $\frac{1}{5}$，正是 (i) 的边界情况。
- (ii) 的答案必须大于 2，因为在 R 上 $2|B| \geq 2$。答案确实如此，而且合理：正态分布在 1 以外的尾部很薄，所以 $|B|$ 通常只比 1 大一点（它的条件平均值约为 1.53）。
- 以 2×10^6 个标准正态的 B 值进行模拟，得出 0.3177、0.4855 和 3.049，与 0.317、0.485 和 3.05 吻合。

常见错误

- 忘记取条件：在 (i) 和 (ii) 中，我们必须除以 $\mathrm{P}(R) = 0.317$。无条件概率 $\Phi(2.6) - \Phi(1) = 0.154$ 并不是答案。
- 在 (i) 中把 $B \geq 1$ 也计算在内。这时两根都是负数，所以没有一个大于 $\frac{1}{5}$。
- 把 $\mathrm{E}|X_1 + X_2|$ 当作 $|\mathrm{E}(X_1 + X_2)|$ 来计算，而后者由对称性可知等于 0。

STEP 之外

$\phi(1)/(1 - \Phi(1))$ 这个量是截尾（truncated）正态分布的平均值，而比值 $\phi(a)/(1 - \Phi(a))$ 称为逆米尔斯比率（inverse Mills ratio）；它出现于统计学（例如设限数据（censored data）的模型），也出现于可靠性理论（reliability theory），在那里它是正态分布的风险率（hazard rate）。关于随机多项式（random polynomials）的问题，例如系数随机的多项式平均有多少个实根，构成了概率论中一个丰富得出人意表的领域。关于条件期望值和连续分布，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门），或 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论导引）。

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。