详细解答

STEP II 1988，第 15 题

概率与统计 • 离散均匀分布变量之和

STEP II 1988，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/88-s2-q15

关键思路

每份试卷的分数在 $\{0, 1, \ldots, m\}$ 上均匀分布（uniform distribution），所以每一组可能的分数出现的机会均等，而概率（probability）就是符合条件的分数组合数目除以所有组合的总数。两份试卷的数目只须简单列举便可得到；三份试卷的数目，则可先以第三份试卷的分数为条件，再把两份试卷的数目相加而得（这就是*卷积*（convolution））。之后利用对称性（symmetry）（把每份试卷的分数作 $k \mapsto m - k$ 的变换），最高的那个范围几乎不费工夫便可得出。

动笔之前先提醒一句：试卷上印出的三份试卷公式有一处印刷错误。下文会推导出正确的公式，并指出印刷版本错在哪里。

解答

设 X_1, X_2, X_3 为各份试卷的分数。由于各份试卷独立评分，它们是独立（independent）的，而且各自在 $\{0, 1, \ldots, m\}$ 上均匀分布。因此每个有序对 (X_1, X_2) 的概率是 $(m+1)^{-2}$，每个有序三元组的概率是 $(m+1)^{-3}$。记 $N_2(n)$ 和 $N_3(n)$ 分别为分数取自 $\{0, \ldots, m\}$ 而总和为 n 的有序对和有序三元组的数目。

两份试卷

我们需要 $X_1 = k$ 及 $X_2 = n - k$，而且两者都在 $[0, m]$ 之内，所以

$$\max(0,\, n - m) \le k \le \min(m,\, n).$$

- 若 $0 \le n \le m$：k 由 0 取到 n，所以 $N_2(n) = n + 1$。

▶ 若 $m < n \le 2m$：k 由 $n-m$ 取到 m，所以 $N_2(n) = m - (n-m) + 1 = 2m - n + 1$。

因此

$$P(X_1 + X_2 = n) = \begin{cases} \dfrac{n+1}{(m+1)^2}, & 0 \le n \le m, \\ \dfrac{2m-n+1}{(m+1)^2}, & m < n \le 2m, \end{cases}$$

这就是题目要求的 $m < n \le 2m$ 的结果，连同 $0 \le n \le m$ 的结果。这个分布形如“帐篷”：每一步上升一，直到 $n = m$，然后逐步下降。

三份试卷

以第三份试卷的分数为条件：若 $X_3 = k$，则需要 $X_1 + X_2 = n - k$，所以

$$N_3(n) = \sum_{k=0}^{m} N_2(n-k) = \sum_{j=n-m}^{n} N_2(j),$$

其中 $j = n - k$，而当 $j < 0$ 或 $j > 2m$ 时 $N_2(j) = 0$。

情况 $0 \le n \le m$。 只有 $j = 0, 1, \ldots, n$ 有贡献（负的 j 给出零），而且它们全都落在帐篷上升的一段：

$$N_3(n) = \sum_{j=0}^{n} (j+1) = \frac{(n+1)(n+2)}{2}, \qquad P(X_1 + X_2 + X_3 = n) = \frac{(n+1)(n+2)}{2(m+1)^3}.$$

情况 $m < n \le 2m$。 这时 $1 \le n - m \le m < n \le 2m$，所以 $j = n-m, \ldots, n$ 这个范围横跨帐篷的顶点。我们在顶点处把它分成两段：

$$N_3(n) = \sum_{j=n-m}^{m} (j+1) + \sum_{j=m+1}^{n} (2m - j + 1).$$

第一个和是把由 $n - m + 1$ 到 $m + 1$ 的整数相加：

$$\sum_{j=n-m}^{m} (j+1) = \frac{(m+1)(m+2) - (n-m)(n-m+1)}{2}.$$

在第二个和中，设 $i = 2m - j + 1$，它由 $2m - n + 1$ 取到 m：

$$\sum_{j=m+1}^{n} (2m - j + 1) = \frac{m(m+1) - (2m-n)(2m-n+1)}{2}.$$

把四个乘积展开，

$$\begin{aligned} 2N_3(n) &= (m^2 + 3m + 2) - (n^2 - 2mn + m^2 + n - m) \\ &\quad + (m^2 + m) - (4m^2 - 4mn + n^2 + 2m - n) \\ &= 6mn - 3m^2 - 2n^2 + 3m + 2. \end{aligned}$$

除以 $(m+1)^3$ 个机会均等的三元组，

$$P(X_1+X_2+X_3=n)=\frac{6mn-3m^2-2n^2+3m+2}{2(m+1)^3},\qquad m<n\le 2m.$$

想一想 这并不是题目印出的公式。究竟谁对？

印刷版本以 $-4m^2$ 代替了 $-3m^2$，又以 $(m+1)^2$ 代替了 $(m+1)^3$；两处都是印刷错误。只须看一个例子便可定夺：取 $m=1$ 及 $n=2$，由 $\{0,1\}$ 取出的三个分数，在 8 种机会均等的情形中恰好有 3 种总和为 2，所以概率是 $\frac{3}{8}$。我们的公式给出 $\frac{12-3-8+3+2}{16}=\frac{3}{8}$，而印刷的公式却给出 $\frac{12-4-8+3+2}{8}=\frac{5}{8}$。无论如何，分母必定是 $(m+1)^3$，因为这正是机会均等的结果的总数。再多验算一次：当 $n=m$ 时，我们中间范围的公式给出 $\frac{(m+1)(m+2)}{2(m+1)^3}$，与第一种情况吻合；这是理所当然的，因为两个和在该处正好相接。

情况 $2m<n\le 3m$。 把每个分数 X_i 换成 $m-X_i$。新的分数同样在 $\{0,\dots,m\}$ 上均匀分布，而三个新分数仍然独立，所以

$$\begin{aligned}P(X_1+X_2+X_3=n)&=P\big((m-X_1)+(m-X_2)+(m-X_3)=n\big)\\&=P(X_1+X_2+X_3=3m-n).\end{aligned}$$

当 $2m<n\le 3m$ 时，有 $0\le 3m-n<m$，所以可把第一种情况中的 n 换成 $3m-n$ 来应用：

$$P(X_1+X_2+X_3=n)=\frac{(3m-n+1)(3m-n+2)}{2(m+1)^3},\qquad 2m<n\le 3m.$$

（同样的对称性，以 $2m-n$ 代替 $3m-n$，可把两份试卷在 $n\le m$ 时的结果变成 $n>m$ 时的结果。）

验算

- 小例子：当 $m=1$ 时，三条公式对 $n=0,1,2,3$ 分别给出 $\frac{1}{8},\frac{3}{8},\frac{3}{8},\frac{1}{8}$，它们的总和是 1，并且与掷三次公平硬币的二项分布（binomial distribution）一致。
- 中间范围的公式在 $n\mapsto 3m-n$ 之下对称：在 $6mn-3m^2-2n^2$ 中把 n 换成 $3m-n$，式子保持不变，正如对称性论证所要求的。
- 对由 1 到 8 的每一个 m 作完全列举，五条公式全部与实际数目完全相符；例如 $m=10$ 及 $n=15$ 时，两种方法都给出 $\frac{91}{1331}$。

常见错误

- 忘记 $N_2(j)$ 有两条不同的公式，在中间范围的求和中由头到尾都用 $j+1$。

▶ 一心想凑出印刷的中间公式，于是“硬砌”代数运算。如果仔细计算的结果与题目给出的答案不符，先用一个极小的例子检验，不要急于怀疑自己。

STEP 之外

独立随机变量（random variable）之和的分布，就是它们各自分布的*卷积*，而我们的和 $\sum_k N_2(n-k)$ 正是如此。概率母函数（probability generating function）可把这个过程变成机械式的运算：数目 $N_3(n)$ 就是 $(1+t+\cdots+t^m)^3$ 中 t^n 的系数。随着试卷的数目愈来愈多，分布的形状会由帐篷变成平滑的隆起，并趋近正态曲线（这就是中心极限定理（central limit theorem））。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）以合适的程度讲解母函数和中心极限定理。

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。