详细解答

STEP III 2000，第 14 题

概率与统计 · 协方差与独立性

STEP III 2000，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/00-s3-q14

关键思路

先补上欠缺的第四个概率（probability），然后把 $\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$ 逐项写出。所有项会合并成单一因子 $a - pq$ 乘以一个不可能为零的量，而 $a = pq$ 正好就是说左上角那一格可以分解成边际概率之积。在 2×2 的表中，只要有一格可以这样分解，边际概率就会迫使其余三格也一样。若变量可取三个值，回旋的空间就多了，一个对称的表便足以打破这种联系。

解答

恒等式

四个概率之和为 1，所以

$$\mathrm{P}(X = x_2,\, Y = y_2) = 1 - a - (q - a) - (p - a) = 1 - p - q + a.$$

沿各行各列相加，便得到边际分布（marginal distribution）：

$$\mathrm{P}(X = x_1) = a + (q - a) = q, \qquad \mathrm{P}(Y = y_1) = a + (p - a) = p,$$

所以 $\mathrm{P}(X = x_2) = 1 - q$ 及 $\mathrm{P}(Y = y_2) = 1 - p$。这就是 p 和 q 的真正意义。

	y_1	y_2	
x_1	a	$q - a$	q
x_2	$p - a$	$1 - p - q + a$	$1 - q$
	p	$1 - p$	

现在

$$\begin{aligned}\mathrm{E}(XY) &= a\,x_1y_1 + (q-a)\,x_1y_2 + (p-a)\,x_2y_1 + (1-p-q+a)\,x_2y_2,\\ \mathrm{E}(X)\mathrm{E}(Y) &= \big(qx_1 + (1-q)x_2\big)\big(py_1 + (1-p)y_2\big)\\ &= pq\,x_1y_1 + q(1-p)\,x_1y_2 + p(1-q)\,x_2y_1 + (1-p)(1-q)\,x_2y_2.\end{aligned}$$

逐个系数相减：

$$\begin{aligned}x_1y_1:&\quad a - pq,\\ x_1y_2:&\quad (q-a) - (q-pq) = -(a-pq),\\ x_2y_1:&\quad (p-a) - (p-pq) = -(a-pq),\\ x_2y_2:&\quad (1-p-q+a) - (1-p-q+pq) = a-pq.\end{aligned}$$

因此

$$\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = (a-pq)(x_1y_1 - x_1y_2 - x_2y_1 + x_2y_2) = (a-pq)(x_1-x_2)(y_1-y_2).$$

所以若 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$，则左方为零，即 $(a-pq)(x_1-x_2)(y_1-y_2) = 0$，这正是所求。

独立性

任何一对各取两个相异值的随机变量（random variable），其联合概率表都具有上述形式，只要选取适当的 a、p、q 便可（替各格命名即可）。若 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$，则上面的乘积为零。由于 $x_1 \neq x_2$ 及 $y_1 \neq y_2$，最后两个因子均不为零，所以 $a = pq$。

> ? **想一想** 我们现在知道有一格可以分解。为甚么这样就能确定其余三格？

因为其余每一格都等于一个边际概率减去一个已知的格：

$$\begin{aligned}&\mathrm{P}(X=x_1, Y=y_2) = q - pq = q(1-p) = \mathrm{P}(X=x_1)\,\mathrm{P}(Y=y_2),\\ &\mathrm{P}(X=x_2, Y=y_1) = p - pq = p(1-q) = \mathrm{P}(X=x_2)\,\mathrm{P}(Y=y_1),\\ &\mathrm{P}(X=x_2, Y=y_2) = 1-p-q+pq = (1-p)(1-q) = \mathrm{P}(X=x_2)\,\mathrm{P}(Y=y_2).\end{aligned}$$

每一格都是其边际概率之积，所以 X 与 Y 独立（independent）。（反过来，独立性可推出 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$，这是标准结果，所以对于只取两个值的变量，这两个条件是等价的。）

取三个值的例子

由于 A 和 B 各自在 $\{-1, 0, 1\}$ 上均匀分布，我们有 $\mathrm{E}(A) = \mathrm{E}(B) = 0$，所以我们需要一个表，其每行每列之和均为 $\frac{1}{3}$，且 $\mathrm{E}(AB) = 0$，但并非每一格都等于 $\frac{1}{9}$。一个对称的选择

便可做到：

	$B=-1$	$B=0$	$B=1$
$A=-1$	$\frac{1}{6}$	0	$\frac{1}{6}$
$A=0$	0	$\frac{1}{3}$	0
$A=1$	$\frac{1}{6}$	0	$\frac{1}{6}$

每行及每列之和均为 $\frac{1}{3}$，所以 A 和 B 各自取 -1、0、1 的概率均为 $\frac{1}{3}$。只有四个角落的格才给出非零的乘积 ab：

$$\mathrm{E}(AB)=\tfrac{1}{6}(1)+\tfrac{1}{6}(-1)+\tfrac{1}{6}(-1)+\tfrac{1}{6}(1)=0=\mathrm{E}(A)\mathrm{E}(B).$$

但 $\mathrm{P}(A=0,B=0)=\frac{1}{3}\neq\frac{1}{9}=\mathrm{P}(A=0)\,\mathrm{P}(B=0)$，所以 A 与 B 并不独立。用文字来说：$B=0$ 当且仅当 $A=0$，否则无论 A 是甚么，B 取 ±1 的机会均等。

验算

- 试取 $X=Y$，其中 $x_1=y_1=1$，$x_2=y_2=0$：此时 $a=p=q$，而恒等式给出 $\mathrm{Var}(X)=p-p^2$，这对伯努利变量（Bernoulli variable）而言是正确的。
- 这个恒等式是一个多项式恒等式：我们随机选取了 2000 组 x_i、y_i、a、p、q，以有理数运算精确地验证了它。同一次验算亦检查了取三个值的表：边际概率全为 $\frac{1}{3}$，$\mathrm{E}(AB)=0$，以及 $\mathrm{P}(A=0,B=0)=\frac{1}{3}$。

常见错误

- 证出 $a=pq$ 便停笔。独立性要求每一格都可以分解，所以其余三格也需要检查（很快便可完成）。
- 忘记说明为何可以约去 $(x_1-x_2)(y_1-y_2)$：因为所取的值相异，所以它不为零。
- 在最后一部分提出 $B=A^2$ 或 $B=|A|$。这些例子确实是不独立而协方差为零，但 B 永不等于 -1，所以它们没有题目要求的边际分布。

STEP 之外

$\mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)$ 这个量就是协方差（covariance），而本题说明“不相关”（uncorrelated）比“独立”为弱，只有取两个值的（类似指示变量的）变量例外：对这类变量，恒等式把一切归结为两个指示变量（indicator）的协方差。你日后会遇到一个相关的事实：对于联合正态（jointly normal）的随机变量，协方差为零确实可推出独立。想进一步了解，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门），或 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论导论）。

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。