详细解答

STEP III 2006，第 14 题

概率与统计 · 协方差与独立性

> **STEP III 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-s3-q14

关键思路

这里的一切都建基于两个事实：期望值(expectation)是线性的；而对于独立(independent)的变量，乘积的期望值等于期望值的乘积。方差(variance)可由 $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2$ 求得，所以我们自始至终只需要计算期望值。要证明两个变量并非独立，只须找出独立性的一个推论不成立便已足够，而最容易检验的就是 $\mathrm{E}(UV) = \mathrm{E}(U)\,\mathrm{E}(V)$。

解答

两个事实

对于任意随机变量（random variable）及常数 a、b，

$$\mathrm{E}(aX_1 + bX_2) = a\,\mathrm{E}(X_1) + b\,\mathrm{E}(X_2),$$

而若 X_1 与 X_2 独立，则

$$\mathrm{E}(X_1X_2) = \mathrm{E}(X_1)\,\mathrm{E}(X_2).$$

第二个事实同样适用于这些变量的函数：若 X_1、X_2 独立，则 X_1^2 与 X_2、X_1 与 X_2^2，以及 X_1^2 与 X_2^2 亦各自独立。我们还会用到 $\mathrm{E}(X_i^2) = \sigma_i^2 + \mu_i^2$，这不过是把 $\mathrm{Var}(X_i) = \mathrm{E}(X_i^2) - \mu_i^2$ 移项而已。

周界与面积

周界为 $P = 2X_1 + 2X_2$，面积为 $A = X_1X_2$。由线性及独立性，

$$\mathrm{E}(P) = 2(\mu_1 + \mu_2), \qquad \mathrm{E}(A) = \mu_1\mu_2.$$

要求 P 的方差，我们从 $\mathrm{E}(P^2)$ 入手：

$$\begin{aligned}\mathrm{E}(P^2) &= 4\,\mathrm{E}(X_1^2 + 2X_1X_2 + X_2^2)\\ &= 4(\sigma_1^2 + \mu_1^2 + 2\mu_1\mu_2 + \sigma_2^2 + \mu_2^2)\\ &= 4(\sigma_1^2 + \sigma_2^2) + 4(\mu_1 + \mu_2)^2,\end{aligned}$$

其中用到线性及 $\mathrm{E}(X_1X_2) = \mu_1\mu_2$。减去 $[\mathrm{E}(P)]^2 = 4(\mu_1 + \mu_2)^2$，得

$$\mathrm{Var}(P) = 4(\sigma_1^2 + \sigma_2^2), \qquad \text{所以}P\text{ 的标准差（standard deviation）为}2\sqrt{\sigma_1^2 + \sigma_2^2}.$$

至于 A，由于 X_1^2 与 X_2^2 独立，

$$\mathrm{E}(A^2) = \mathrm{E}(X_1^2)\,\mathrm{E}(X_2^2) = (\sigma_1^2 + \mu_1^2)(\sigma_2^2 + \mu_2^2),$$

减去 $\mu_1^2\mu_2^2$ 后余下

$$\mathrm{Var}(A) = \sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2,$$

所以 A 的标准差为

$$\sqrt{\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2}.$$

P 与 A 并非独立

假如 P 与 A 独立，便会有 $\mathrm{E}(PA) = \mathrm{E}(P)\,\mathrm{E}(A)$。现在 $PA = 2X_1^2X_2 + 2X_1X_2^2$，所以由线性以及 X_1 与 X_2 的独立性，

$$\begin{aligned}\mathrm{E}(PA) &= 2(\sigma_1^2 + \mu_1^2)\mu_2 + 2\mu_1(\sigma_2^2 + \mu_2^2)\\ &= 2(\mu_1 + \mu_2)\mu_1\mu_2 + 2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2)\\ &= \mathrm{E}(P)\,\mathrm{E}(A) + 2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2).\end{aligned}$$

长度和阔度都是正数, 所以 $\mu_1, \mu_2 > 0$, 而且 $\sigma_1, \sigma_2 \neq 0$, 因此多出来的一项 $2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2)$ 严格为正。故此 $\mathrm{E}(PA) \neq \mathrm{E}(P)\,\mathrm{E}(A)$，即 P 与 A 并非独立。

变量 $Z = P - \alpha A$

同样地，若 Z 与 A 独立，则必有 $\mathrm{E}(ZA) = \mathrm{E}(Z)\,\mathrm{E}(A)$。由线性，

$$\begin{aligned}\mathrm{E}(ZA) - \mathrm{E}(Z)\,\mathrm{E}(A) &= [\mathrm{E}(PA) - \mathrm{E}(P)\,\mathrm{E}(A)] - \alpha[\mathrm{E}(A^2) - [\mathrm{E}(A)]^2]\\ &= 2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2) - \alpha(\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2),\end{aligned}$$

其中用到上面的两个结果。与 α 相乘的括号就是 $\mathrm{Var}(A)$, 它是正数 (当中含有 $\sigma_1^2\sigma_2^2 > 0$), 所以这个式子恰好只在 α 的一个值处等于零：

$$\alpha_0 = \frac{2(\mu_1\sigma_2^2 + \mu_2\sigma_1^2)}{\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2}.$$

若 $\alpha \neq \alpha_0$，则 $\mathrm{E}(ZA) \neq \mathrm{E}(Z)\,\mathrm{E}(A)$，所以 Z 与 A 并非独立，这正是所要证明的。

想一想 $\alpha = \alpha_0$ 是否表示 Z 与 A 确实独立?

不是。协方差（covariance）为零是独立性的推论，却不能保证独立。当 $\alpha = \alpha_0$ 时，我们的检验根本提供不了任何信息，必须另觅论证。最后一部分说明，在一个具体例子中，独立性在该处同样不成立。

最后一部分

这里 $\mu_1 = \mu_2 = 2$ 而 $\sigma_1^2 = \sigma_2^2 = 1$，所以

$$\alpha_0 = \frac{2(2+2)}{4+4+1} = \frac{8}{9}.$$

由上一部分，对任何 $\alpha \neq \frac{8}{9}$，Z 与 A 都并非独立。若要用一个论证涵盖所有 α，可列出四个等可能的结果：

(X_1, X_2)	A	P	$Z = P - \alpha A$
$(1,1)$	1	4	$4-\alpha$
$(1,3),(3,1)$	3	8	$8-3\alpha$
$(3,3)$	9	12	$12-9\alpha$

关键的观察是：P 完全由 A 决定，因而 Z 也是如此。设 $z = 4-\alpha$。知道 $A = 1$ 便必有 $Z = z$，所以

$$\mathrm{P}(Z = z,\ A = 1) = \mathrm{P}(A = 1) = \tfrac{1}{4}.$$

若要独立，这个概率便须等于 $\mathrm{P}(Z = z)\,\mathrm{P}(A = 1)$，即 $\mathrm{P}(Z = z) = 1$，于是 Z 在每一个结果上都要取值 $4-\alpha$。这需要 $8-3\alpha = 4-\alpha$，即 $\alpha = 2$，同时又需要 $12-9\alpha = 4-\alpha$，即 $\alpha = 1$。两者不可能同时成立，所以 $\mathrm{P}(Z = z) < 1$，独立性不成立。

因此，无论 α 取甚么值，Z 与 A 都并非独立。例如当 $\alpha = \frac{8}{9}$ 时，Z 的取值为 $\frac{28}{9}$、$\frac{16}{3}$ 和 4，而 $\mathrm{P}(Z = \frac{28}{9}, A = 1) = \frac{1}{4}$，但 $\mathrm{P}(Z = \frac{28}{9})\,\mathrm{P}(A = 1) = \frac{1}{16}$。

验算

- 若完全没有变异（$\sigma_1 = \sigma_2 = 0$），两个标准差都化为 0，理应如此。
- 在最后一部分，按公式得 $\mathrm{Var}(A) = 1+4+4 = 9$；直接计算，$A$ 取值 $1, 3, 9$ 的概率分别为 $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$，得出 $\mathrm{E}(A) = 4$、$\mathrm{E}(A^2) = 25$，方差为 9。
- 我们以分数作精确计算，在 200 个随机的离散分布上验证了平均值、方差和协方差的一般公式。我们亦确认了在 α_0 处 $\mathrm{E}(ZA) = \mathrm{E}(Z)\mathrm{E}(A)$，确认了最后一部分中 $\alpha_0 = \frac{8}{9}$，并确认了在该部分中，对于 $\alpha = \frac{8}{9}$ 及另外 500 个 α 值，Z 与 A 均不独立。

常见错误

- 写成 $\mathrm{E}(X_1^2) = \mu_1^2$。正是方差项 σ_1^2 令 $\mathrm{Var}(A)$ 和协方差不为零。
- 当 $\alpha = \alpha_0$ 时断定 Z 与 A 独立。协方差为零并不蕴涵独立。
- 忘记说明为何 $\mu_2\sigma_1^2 + \mu_1\sigma_2^2 \neq 0$。正是因为长度为正数，各项才不可能互相抵销。

STEP 之外

$\mathrm{E}(UV) - \mathrm{E}(U)\mathrm{E}(V)$ 这个量就是协方差，而特殊值 α_0 等于 $\mathrm{Cov}(P, A)/\mathrm{Var}(A)$，即 P 对 A 的最小二乘回归（least-squares regression）直线的斜率。取 $\alpha = \alpha_0$ 会令残差（residual）Z 与 A 不相关（uncorrelated），但并不令两者独立。对于联合正态（jointly normal）的变量，这两个概念是一致的，这也是正态分布如此方便的原因之一。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门）对协方差、相关与独立性有细致的论述。

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。