

On the statistical independence in a contingency table

A. H. JOARDER

School of Mathematics and Statistics, University of Sydney, NSW, Australia 2006

(Received 18 December 1996)

In this pedagogical note, a necessary and sufficient condition for the independence of two elementary events in a 2×2 contingency table is discussed for the purpose of making the classroom teaching of these notions more expository and informative. This in turn shows equivalence of statistical independence of events and linear dependence of rows or columns in a contingency table.

1. Introduction

Suppose that a population has been cross-classified according to two attributes each with two categories. It is well-known that two non-null events are independent if the conditional probability of the occurrence of A given that B has occurred is the same as the unconditional probability of the occurrence of A , i.e. $P(A | B) = P(A)$ (see [1], p. 40). This rule has been used for checking the independence of elementary events in a 2×2 contingency table obtained by decomposing a population into two attributes each having two categories.

2. Main points

We consider the problem of statistical independence of two elementary events defined in a 2×2 contingency table. A condition for the independence of two events whenever the population under investigation can be cross-classified in a 2×2 contingency table is presented below.

Suppose that a population of $n = n_{11} + n_{12} + n_{21} + n_{22}$ individuals is cross-classified according to the attributes A (in rows) and B (in columns) each having 2 categories as shown in the following matrix

	B_1	B_2
A_1	n_{11}	n_{12}
A_2	n_{21}	n_{22}

Then the events A_i and B_j ($i, j = 1, 2$) are independent if and only if $|N| = 0$ where A_i = event that an individual belongs to category i ($i = 1$ or 2) of A , B_j = event that an individual belongs to category j ($j = 1$ or 2) of B , and N is the incidence matrix given by

$$N = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix}$$

The above result is obvious but interesting. By definition, each pair of events

$$(i) A_1 \text{ and } B_1 \quad (ii) A_1 \text{ and } B_2 \quad (iii) A_2 \text{ and } B_1 \text{ and } (iv) A_2 \text{ and } B_2 \quad (1)$$

are independent if

$$(i) P(A_1 | B_1) = P(A_1) \quad (ii) P(A_2 | B_1) = P(A_2),$$

$$(iii) P(A_1 | B_2) = P(A_1) \quad (iv) P(A_2 | B_2) = P(A_2)$$

i.e. if

$$(i) \frac{n_{11}}{n_{11} + n_{21}} = \frac{n_{11} + n_{12}}{n} \quad (ii) \frac{n_{21}}{n_{11} + n_{21}} = \frac{n_{21} + n_{22}}{n},$$

$$(iii) \frac{n_{12}}{n_{12} + n_{22}} = \frac{n_{11} + n_{12}}{n} \quad (iv) \frac{n_{22}}{n_{12} + n_{22}} = \frac{n_{21} + n_{22}}{n} \quad (2)$$

i.e. if

$$n_{11}n_{22} - n_{21}n_{12} = 0$$

Thus it is verified that each pair of events (1) are independent if $|N| = 0$.

On the other hand, if $|N| = 0$, then $n_{11}n_{22} = n_{21}n_{12}$. Now by adding some terms in both sides of this equation, we can easily get back to the conditions stated in relations (2). However, under the conditions (2), each pair of events (1) are independent. Hence it is verified that the pairs of events (1) are statistically independent if and only if the determinant of the incidence matrix vanishes, i.e. the rows or the columns are linearly dependent.

The result is now illustrated with an example. Suppose that a proficiency test for a certain skill was given to 100 employees of a firm. The breakdown of sex (male and female in columns) and proficiency (pass and fail in rows) is summarized in the following matrix

	Male	Female
Pass	30	10
Fail	45	15

Here $P(\text{pass} | \text{female}) = 10/25$ and $P(\text{pass}) = 40/100$. Since the two probabilities are the same the events 'pass' and 'female' are statistically independent. Here the determinant of the incidence matrix vanishes. Thus the events A_i ($i = 1$ or 2) and B_j ($j = 1$ or 2) are statistically independent if and only if the rows or the columns of the contingency table are linearly dependent.

The above idea is also applicable to an $r \times c$ contingency table collapsed into an appropriate 2×2 contingency table. Consider the following hypothetical example of 160 students classified according to intelligence and dress:

	Dull	Intelligent	Very capable
Poorly clothed	18	20	26
Well clothed	17	16	30
Very well clothed	10	14	09

Thus the events 'dull' and 'poorly clothed' are independent since the rows of the following matrix are linearly dependent:

	Dull	Not dull
Poorly clothed	18	46
Not poorly clothed	27	69

while the events 'very capable' and 'very well clothed' are not independent since the rows of the following matrix are not linearly dependent:

	Not very capable	Very capable
Not very well clothed	71	56
Very well clothed	24	09

Finally, we note that the above result also applied to any $r \times c$ contingency table collapsed into an appropriate $r \times r$ contingency table. This means that the events A_i ($i = 1$ or 2 or $\dots$ or r) and B_j ($j = 1$ or 2 or $\dots$ or c) are statistically independent if and only if the rows or the columns of the contingency table are linearly dependent.

Reference

- [1] MOOD, A. M., GRAYBILL, F. A., and BOES, D. C., 1974, *Introduction to the Theory of Statistics*, 3rd edn (New York: McGraw-Hill).

A path having all of its subpaths of infinite length

MARK LYNCH

Department of Mathematics, Millsaps College, Jackson, MS 39210, USA
email: Lynchmj@Millsaps.edu

(Received 2 January 1997)

We give an example of a continuous function $f : [0, 1] \rightarrow \mathbb{R}$ such that the length of the graph of $f(x)$ restricted to any subinterval of $[0, 1]$ has infinite length. The construction uses elementary topological concepts of the real plane normally covered in the first semester of undergraduate analysis and topology courses and does not involve uniform convergence.

Classroom notes

An interesting property of pythagorean triplets

M. N. DESHPANDE

Institute of Science, Nagpur - 440 001 India

(Received 8 February 1996)

1. Introduction

A *Pythagorean triplet* is a set of three numbers a , b and c such that $0 < a < b < c$ and $a^2 + b^2 = c^2$.

Here we establish an interesting property of Pythagorean triplets which shows that six squares are associated with each Pythagorean triplet.

Property. For each Pythagorean triplet (a, b, c) , there exist two positive integers k and d such that the following six expressions are perfect squares:

$$\begin{array}{lll} (1) \ ab - k & (2) \ ac - k & (3) \ bc - k \\ (4) \ k^2 - ad & (5) \ k^2 - bd & (6) \ k^2 - cd \end{array}$$

Before proving the property in the following section, we give below a few illustrations.

Pythagorean triplets			k	d
3	4	5	11	24
5	12	13	56	240
20	21	29	384	5 040
48	55	73	2 415	79 200
85	132	157	10 320	673 200

Proof of the property. We denote as usual by s , the semiperimeter, i.e. $s = (a + b + c)/2$.

Consider $k = s^2 - c^2$ and $d = 4s(s - c)^2$. It is well known that, for Pythagorean triplets, $a + b + c$ is always even and hence k and d are positive integers.

$$\begin{aligned} (1) \ ab - k &= [(a + b)^2 - a^2 - b^2]/2 - s^2 + c^2 \\ &= [(2s - c)^2 - c^2]/2 - s^2 + c^2 \\ &= 2s^2 - 2sc - s^2 + c^2 \\ &= (s - c)^2 \end{aligned}$$

Similarly we can show that