The dependence structure of conditional probabilities in a contingency table

Anwar H. Joarder and Walid S. Al-Sabah

Department of Mathematical Sciences
King Fahd University of Petroleum & Minerals
Dhahran 312361, Saudi Arabia
Emails: anwarj@kfupm.edu.sa and walid@kfupm.edu.sa

Conditional probability and statistical independence can better be explained with contingency tables. In this note some special cases of 2×2 contingency table is considered. In turn an interesting insight into statistical dependence as well as independence of events is obtained.

***Keywords**: Conditional probability; contingency table; incidence matrix; singularity; statistical independence*

1. Introduction

Elementary probabilities are obtained for the outcomes of situations conveniently called random experiments. They are usually taught with the help of examples of dice, coins and cards. Not everybody feels comfortable with these approaches. Experience shows that
conditional probability and statistical independence can better be explained with contingency tables often encountered by them in real life. Consider a general 2×2 contingency table

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

The matrix given by

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

will hereinafter be called incidence matrix. In this note some special cases of 2×2 contingency table is considered. In turn a relation is observed between the

dependence structure of conditional probabilities, nonsingularity of the incidence matrix N formed by the square contingency table, and statistical dependence of events. The properties that are going to be discussed here will also be true for any $r \times c$ contingency table collapsed as a 2×2 contingency table.
The notion of statistical independence is closely related to conditional probability. Given that B happens, the probability is

$$\frac{P(A \cap B)}{P(B)}$$

that the event A happens. The above ratio is usually denoted by $P(A \mid B)$ i.e.

$$\frac{P(A \cap B)}{P(B)} = P(A \mid B)\,. \tag{1.1}$$

The left hand side of (1.1) should be emphasized to the students as the right hand side is usually misunderstood by them. If the ratio is the same as $P(A)$, it implies that B does not affect the occurrence of A. In other word, A is statistically independent of B. Thus in this case it follows from (1.1) that

$$P(A \cap B) = P(A)P(B) \tag{1.2}$$

which is used as the definition of statistical independence in many books. It follows from (1.2) that if A is statistically independent of B, then B is statistically independent of A.

Consider the independence of the categories of two attributes A and B. By definition each pair of events (i) A_1 and B_1 (ii) A_1 and B_2 (iii) A_2 and B_1 and (iv) A_2 and B_2 are independent if the following conditions hold:

$$\begin{aligned} &(i)\ P(A_1 \mid B_1) = P(A_1) \\ &(ii)\ P(A_1 \mid B_2) = P(A_1) \\ &(iii)\ P(A_2 \mid B_1) = P(A_2) \text{ and} \\ &(iv)\ P(A_2 \mid B_2) = P(A_2) \end{aligned} \tag{1.3}$$

respectively. But it is straightforward to prove that the above four $(= 2^2)$ conditions are equivalent (Hines and Montgomery, 1990, p.51). Thus if A_1 and B_1 are independent, then so are (a) A_1 and B_2 , (b) A_2 and B_1 and (c) A_2 and B_2. That is if any pair of events are independent in a 2×2 table, then other three pair of events in (1.3) are independent and not mutually exclusive.

In what follows we provide two interesting results that provide some insight into statistical independence. They follow from the rearrangement of the equations in (1.3).

(1) For any contingency table having attributes A and B with categories $A_1,\ A_2$ and categories $B_1,\ B_2$ respectively, the events A_1 and B_1 are independent *if and only* if $A \mid B_1$ and $A \mid B_2$ have the same probability distribution i.e.

$$\begin{aligned} &(i) \quad P(A_1 \mid B_1) = P(A_1 \mid B_2) \\ &(ii) \quad P(A_2 \mid B_1) = P(A_2 \mid B_2) \end{aligned} \tag{1.4}$$

In a 2×2 contingency table it is conventional to write $A_1 = A$ and $B_1 = B$ so that $A_2 = \overline{A}$ and $B_2 = \overline{B}$. To explain (1.4), consider the following example of the breakdown of computers having circuit boards for a modem (A) or for a printer (B):

	A	$\overline{A}$	
B	10	15	25
$\overline{B}$	30	45	75
	40	60	100

The events A and B are independent *if and only* if $A \mid B$ and $A \mid \overline{B}$ have the same probability distribution i.e.

$(i) \quad P(A \mid B),\ P(\overline{A} \mid B)$ and

$(ii) \quad P(A \mid \overline{B}),\ P(\overline{A} \mid \overline{B})$

are the same. Since the two sets of probabilities

$$(i) \quad P(A \mid B) = \frac{10}{25} = 0.40,\quad P(\overline{A} \mid B) = \frac{15}{25} = 0.60 \quad \text{and}$$

$$(ii) \quad P(A \mid \overline{B}) = \frac{30}{75} = 0.40,\quad P(\overline{A} \mid \overline{B}) = \frac{45}{75} = 0.60$$

are the same, the events A and B are independent .

(2) For any contingency table having attributes A and B with categories $A_1,\ A_2$ and categories $B_1,\ B_2$ respectively, the events A_1 and B_1 are independent if and only

$$\begin{aligned} &(i)\ P(A_1 \mid B_1) = P(A_1 \mid B_2) = P(A_1) \text{ and} \\ &(ii)\ P(A_2 \mid B_1) = P(A_2 \mid B_2) = P(A_2) \end{aligned} \tag{1.5}$$

The equation (i) of (1.5) says that neither the occurrence of B_1 nor B_2 affects the occurrence of A_1. Similarly the equation *(ii)* of (1.5) indicates that neither the occurrence of B_1 nor B_2 affects the occurrence of A_2.

In what follows we provide two other interesting results that are special cases of a 2×2 contingency table:

(1) For any contingency table having attributes A and B with categories $A_1,\ A_2$ and categories $B_1,\ B_2$ respectively, the following holds:

$P(A_1 \cap B_1) = P(A_2 \cap B_2)$ if and only if $P(A_1) = P(A_2),\ P(B_1) = P(B_2)$.

This means that the 2×2 incidence matrix has equal diagonal elements.

(2) For any contingency table having attributes A and B with categories $A_1,\ A_2$ and categories $B_1,\ B_2$ respectively, the following holds:

$$\frac{P(A_1)}{P(A_2)} = \frac{P(B_1)}{P(B_2)}$$

if and only if $P(A_1 \cap B_1) + P(A_2 \cap B_2) = P(A_1 \cap B_2) + P(A_2 \cap B_1)$.

This implies that the sum of the diagonal elements is the same as that of the off-diagonal elements. Thus the probability of having exactly one of the two attributes is the same as having none or both the attributes.

2. The Main Result

The main result is presented below in the form of a theorem.

Theorem 2.1 For any contingency table having attributes A and B with categories $A_1,\ A_2$ and $B_1,\ B_2$ respectively, the incidence matrix has the following implications:

(a) $P(A_1 \mid B_1) < P(A_1) < P(A_1 \mid B_2) \text{ iff } |N| < 0$ (2.1)

(b) $P(A_1 \mid B_1) = P(A_1) = P(A_1 \mid B_2) \text{ iff } |N| = 0$ (2.2)

(c) $P(A_1 \mid B_1) > P(A_1) > P(A_1 \mid B_2) \text{ iff } |N| > 0$ (2.3)

Proof: *(a)* Let $P(A_1 \mid B_1) < P(A_1) < P(A_1 \mid B_2)$. Then

$$\frac{n_{11}}{n_{11}+n_{21}} < \frac{n_{11}+n_{12}}{n} \text{ and } \frac{n_{11}+n_{12}}{n} < \frac{n_{12}}{n_{12}+n_{22}} .$$

Writing out $n = n_{11} + n_{12} + n_{21} + n_{22}$ and simplifying, we have from each of the inequality

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

or $n_{11}n_{22} < n_{12}n_{21}$ (i.e. $|N| < 0$). (2.4)

Again let $|N| < 0$, i.e. $n_{11}n_{22} < n_{12}n_{21}$. Now adding $n_{11}(n_{11} + n_{12} + n_{21})$ to both sides of this inequality, we have

$n_{11}n_{22} + n_{11}(n_{11} + n_{12} + n_{21}) < n_{12}n_{21} + n_{11}(n_{11} + n_{12} + n_{21})$

i.e. $n_{11}\, n < (n_{11} + n_{12})(n_{11} + n_{21})$.

Dividing both sides by $n(n_{11} + n_{21})$, we have

$$\frac{n_{11}}{n_{11}+n_{21}} < \frac{n_{11}+n_{12}}{n}, \quad \text{i.e. } P(A_1 \mid B_1) < P(A_1) .$$

Similarly by adding $n_{12}(n_{11} + n_{12} + n_{22})$ to both sides of (2.4), we have

$n_{11}n_{22} + n_{12}(n_{11} + n_{12} + n_{22}) < n_{12}n_{21} + n_{12}(n_{11} + n_{12} + n_{22})$

or, $(n_{11} + n_{12})(n_{12} + n_{22}) < n_{12}(n_{11} + n_{12} + n_{21} + n_{22})$

or, $(n_{11} + n_{12})(n_{12} + n_{22}) < n\, n_{12}$.

Dividing both sides of the resulting inequality by $n(n_{12} + n_{22})$, we have

$$\frac{n_{11}+n_{12}}{n} < \frac{n_{12}}{n_{12}+n_{22}}, \quad \text{i.e. } P(A_1) < P(A_1 \mid B_2) .$$

(b) See Joarder (1998).

(c) The proof is similar to that in part (a) above.

The result in *(a)* here means that A_1 is less likely to happen if B_1 happens, while A_1 is more likely to happen if B_1 does not happen. The result in *(c)* similarly means that A_1 is more likely to happen if B_1 happens, while A_1 is less likely to happen if B_1 does not happen. The result in *(b)* means that the occurrence of B_1 does not affect the occurrence of A_1 and vice versa.

Part *(b)* implies that the events A_1 and B_1 are independent *if and only if* any of the following equivalent conditions is satisfied:

(i) rows are linearly dependent
(ii) columns are linearly dependent
(iii) the incidence matrix N is singular
(iv) $n_{ij} = \dfrac{n_{i.}n_{.j}}{n}$ where $n_{i.} = n_{i1} + n_{i2}$ and $n_{.j} = n_{1j} + n_{2j}$ $(i = 1, 2;\ j = 1, 2)$.

3. Some Illustrations

As earlier let $A_1 = A$ and $B_1 = B$ so that $A_2 = \overline{A}$ and $B_2 = \overline{B}$. To explain *(a)* of Theorem 2.1, consider the following the breakdown of a computer having modem boards (A) or printer boards (B):

	A	$\overline{A}$	
B	4	16	20
$\overline{B}$	36	44	80
	40	60	100

Here the following three probabilities

$$P(A \mid B) = \frac{4}{20} = 0.20,\ P(A) = \frac{40}{100} = 0.40,\ P(A \mid \overline{B}) = \frac{36}{80} = 0.45$$

are not the same. Observe that $|N| < 0$ and $P(A \mid B) < P(A) < P(A \mid \overline{B})$. This means that that computers without printer boards are more likely to have modem boards than computers with printer boards. In other words, they are statistically dependent.

Similarly, the probabilities

$$P(B \mid A) = \frac{4}{40} = 0.10,\ P(B) = \frac{20}{100} = 0.20,\ P(B \mid \overline{A}) = \frac{16}{60} \approx 0.26$$

are not the same. Observe that $|N| < 0$ and $P(B \mid A) < P(B) < P(B \mid \overline{A})$. This means that that computers without modem boards are more likely to have printer boards than computers with modem boards. In other words, they are statistically dependent.

To explain *(c)* of Theorem 2.1, consider the following the breakdown of a computer having a modem board (A) or a circuit board (B):

	A	$\overline{A}$	
B	12	8	20
$\overline{B}$	28	52	80
	40	60	100

Here the following three probabilities

$$P(A \mid B) = \frac{12}{20} = 0.60,\ P(A) = \frac{40}{100} = 0.40,\ P(A \mid \overline{B}) = \frac{28}{80} = 0.35$$

are not the same. Observe that $|N| > 0$ and $P(A \mid B) > P(A) > P(A \mid \overline{B})$. This means that that computers with printer boards are more likely to have modem boards than computers without printer boards. In other words, they are statistically dependent. Similarly, the following three probabilities

$$P(B \mid A) = \frac{12}{40} = 0.30,\ P(B) = \frac{20}{100} = 0.20,\ P(B \mid \overline{A}) = \frac{8}{60} \approx 0.13,$$

are not the same. Observe that $|N| > 0$ and $P(B \mid A) > P(B) > P(B \mid \overline{A})$. This means that that computers with modem boards are more likely to have printer boards than computers without modem boards. In other words, they are statistically dependent.

To explain *(b)* of Theorem 2.1, consider the following the breakdown of a computer having a modem board (A) or a circuit board (B):

	A	$\overline{A}$	
B	10	15	25
$\overline{B}$	30	45	75
	40	60	100

Here the following three probabilities

$$P(A \mid B) = \frac{10}{25} = 0.40,\ P(A) = \frac{40}{100} = 0.40,\ P(A \mid \overline{B}) = \frac{30}{75} = 0.40$$

are the same. Observe that $|N| = 0$ and $P(A \mid B) = P(A) = P(A \mid \overline{B})$. The same is true for the following three probabilities:

$$P(B \mid A) = \frac{10}{40} = 0.25,\ P(B) = \frac{2}{100} = 0.25,\ P(B \mid \overline{A}) = \frac{15}{60} = 0.25,$$

Observe that $|N| = 0$ and $P(B \mid A) = P(B) = P(B \mid \overline{A})$. Since the above three probabilities are the same, it follows that having a modem has nothing to do with having a printer or vice versa. In other words, they are statistically independent.

The notions discussed here are also true for any $r \times c$ contingency table collapsed into an appropriate 2×2 contingency table with categories of interest. We remark that though Theorem 2.1 is proved in the context of contingency table, it is true for any two events

A and B where $A_1 = A, B_1 = B$ so that $A_2 = \overline{A}$, $B_2 = \overline{B}$ and

$$N = \begin{pmatrix} P(AB) & P(A\overline{B}) \\ P(\overline{A}B) & P(\overline{A}\overline{B}) \end{pmatrix}.$$

Acknowledgements

The authors acknowledge the excellent research facilities available at King Fahd University of Petroleum and Minerals, Saudi Arabia.

References

[1] Hines, W.W. and Montgomery, D.C., 1990, *Probability and Statistics in Engineering and Management Sciences*. John Wiley and Sons, New York.

[2] Joarder, A.H., 1998, On the statistical independence in a contingency table. *International Journal of Mathematical Education in Science and Technology*. **29**, No.5, 780-782

File: c\ p61ijmesta.doc (ijmest 01/346)

Theorem: $0 < l, m, r < 1$. Suppose $P(A) = l + m$ and $P(B) = m + r$. Then A and B are independent iff

$$m^2 + (l + r - 1)m + lr = 0$$

Theorem: $0 < l, m, r$. Then A with $l + m$ elements and B with $m + r$ elements are independent iff

$$m^2 + (l + r - n)m + lr = 0$$

where n is the number of elements in the universal set.

Let us check statistical independence by (3.12).

Here $LHS = P(A \cap B) = \frac{10}{100} = 0.10$ and $RHS = P(A)P(B) = \frac{40}{100}\frac{25}{100} = 0.10$. Since the two sides are the same we say A and B are statistically independent.

Let us check statistical independence by (3.12).

Here $LHS = P(A \cap B) = \frac{4}{100}$ and $RHS = P(A)P(B) = \frac{40}{100}\frac{20}{100}$. Since the two sides are not the same we say A and B are statistically dependent.

It is interesting to check that the two equations of (3.1) hold here.

$$P(A \cap B) = \begin{cases} P(A)P(B \mid A) = \frac{40}{100}\frac{4}{40} = \frac{4}{100} \\ P(B)P(A \mid B) = \frac{20}{100}\frac{4}{20} = \frac{4}{100} \end{cases}$$

It is interesting to check that the two equations of (3.1) hold here.

$$P(A \cap B) = \begin{cases} P(A)P(B \mid A) = \frac{40}{100}\frac{10}{40} = \frac{10}{100} \\ P(B)P(A \mid B) = \frac{25}{100}\frac{10}{25} = \frac{10}{100} \end{cases}$$

It may be reminded that in many real world situations only one expression of (3.1) makes sense.

If A and B are independent then the formula in (3.5) can be written as

$$P(A \cup B) = 1 - P(\overline{A \cup B}) = 1 - P(\overline{A})\,P(\overline{B}).$$

Though the above arguments provides insight into the problem, it is often easy to check statistical independence by (3.12).

Here $LHS = P(A \cap B) = \frac{12}{100}$ and $RHS = P(A)P(B) = \frac{40}{100}\frac{20}{100}$. Since the two sides are not the same we say A and B are statistically dependent.

It is interesting to check that the two equations of (3.1) hold here.

$$P(A \cap B) = \begin{cases} P(A)P(B \mid A) = \frac{40}{100}\frac{12}{40} = \frac{12}{100} \\ P(B)P(A \mid B) = \frac{20}{100}\frac{12}{20} = \frac{12}{100} \end{cases}$$

(i) The rows (equivalently the columns) of the incidence matrix $N = (n_{ij}),\ i, j = 1, 2$ are linearly dependent (Joarder, 1998).

(ii) The incidence matrix is singular.

$(iii)\ n_{ij} = \frac{n_{i.}n_{.j}}{n}$

It is interesting to note that if categories are equally likely i.e. $P(A_1) = P(B_1) = \frac{1}{2}$ then A and B will be independent.

Example 3.13 We now provide an example which is to identify managerial prospects as to who are both talented and motivated. A personnel manager constructed the table shown here to define nine combinations of talent-motivation levels.

		TALENT (B)		
		High	Medium	Low
	High	16	20	30
MOTIVATION	Medium	24	30	45
(A)	Low	08	10	15
		48	60	
90				

We notice that the following three sets of probabilities

$P(A \mid B_1):\ \frac{16}{48}, \frac{24}{48}, \frac{8}{48}$

$P(A \mid B_2) : \frac{20}{60}, \frac{30}{60}, \frac{10}{60}$

$P(A \mid B_3) : \frac{30}{90}, \frac{45}{90}, \frac{15}{90}$

are identical which implies the independence of the categories of the attributes A and B.

Finally we comment that categorical variables A and B are independent if and only if the following equivalent conditions hold:

(i) rows are linearly dependent.

(ii) columns are linearly dependent.

(iii) $n_{ij} = \dfrac{n_{i.} n_{.j}}{n}$.

(iv) the incidence matrix N is singular.

(v) $P(A \mid B_1), P(A \mid B_2), \ldots, P(A \mid B_c)$ are identically distributed.

(vi) $P(B \mid A_1), P(B \mid A_2), \ldots, P(B \mid A_r)$ are identically distributed.

(3.13)

Categories $i\,(=1, 2, \ldots, r)$ of A and $j\,(=1, 2, \ldots, c)$ of B are independent if the 2×2 matrix in the following table is singular:

	j	j'
i	n_{ij}	$n_{i.} - n_{ij}$
i'	$n_{.j} - n_{ij}$	$n_{..} - n_{i.} - n_{.j} + n_{ij}$

where category i' means all categories except the i th category.

Example 3.14 Consider the hypothetical example of comparing the fidelity (accuracy of the reproduction of sound) and selectivity of 160 radio receivers. The radio receivers are classified as Low, Medium and High in each of the two attributes.

	Sensitivity Low (B_1)	 Medium (B_2)	 High (B_3)	
Fidelity Low (A_1)	18	20	26	63
Medium (A_2)	17	16	30	64
	10	14	09	33

High (A_3)				
	45	50	65	160

The categories A_2 and B_1 are dependent since the following three probabilities are not the same (see condition *(v)* of (3.13)).

$$P(A_2 \mid B_1) = \frac{17}{45} \approx 0.38$$
$$P(A_2 \mid \overline{B}_1) = \frac{46}{115} = 0.40$$

$$P(A_2) = \frac{63}{160} \approx 0.39$$

which is obvious from the non-singularity (see condition (*iv*) of (3.13)) of the following matrix :

	B_1	$\overline{B}_1$
A_2	17	$16+30=46$
$\overline{A}_2$	$18+10=28$	$(20+26)+(14+9)=69$

However the events A_1 and B_1 are independent since

$$P(A_1 \mid B_1) = \frac{18}{45} = 0.40$$
$$P(A_1 \mid \overline{B}_1) = \frac{20+26}{115} = 0.40$$
$$P(A_1) = \frac{64}{160} = 0.40$$

which is obvious by inspection from the following table (rows are linearly dependent)

	B_1	$\overline{B}_1$	
A_1	18	20+26	64
$\overline{A}_1$	17+10	(16+30)+(14+09)	96
	45	115	160