The dependence structure of conditional probabilities in a contingency table

Anwar H. Joarder, Walid S. Al-Sabah
The dependence structure of conditional probabilities in a contingency table

ANWAR H. JOARDER and WALID S. AL-SABAHH

Department of Mathematical Sciences, King Fahd University of Petroleum and Minerals, Dhahran 312361, Saudi Arabia

Emails: anwarj@kfupm.edu.sa and walid@kfupm.edu.sa

Abstract: Conditional probability and statistical independence can better be explained with contingency tables. In this note some special cases of $2 \times 2$ 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 $2 \times 2$ contingency table

\[
\begin{array}{c|cc}
 & B_1 & B_2 \\
\hline
A_1 & n_{11} & n_{12} \\
A_2 & n_{21} & n_{22} \\
\end{array}
\]

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 $2 \times 2$ 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 $2 \times 2$ 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:

$$(i) \ P(A_1 \mid B_1) = P(A_1) \nonumber$ \\
$$(ii) \ P(A_1 \mid B_2) = P(A_1) \nonumber$ \\
$$(iii) \ P(A_2 \mid B_1) = P(A_2) \text{ and} \nonumber$ \\
$$(iv) \ P(A_2 \mid B_2) = P(A_2) \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 $2 \times 2$ 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 distributions i.e.
(i) \( P(A_1 \mid B_1), \ P(A_2 \mid B_1) \) and

(ii) \( P(A_1 \mid B_2), \ P(A_2 \mid B_2) \)

are the same.

In a 2\( \times \)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 a
computers having circuit boards for a modem (\( A \)) or for a printer (\( B \)):

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>( \overline{A} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>( \overline{B} )</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

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

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

are the same. Since the two sets of probabilities

\[
\begin{align*}
(i) \quad P(A \mid B) &= \frac{10}{25} = 0.40, \quad P(\overline{A} \mid B) = \frac{15}{25} = 0.60 \quad \text{and} \\
(ii) \quad P(A \mid \overline{B}) &= \frac{30}{75} = 0.40, \quad P(\overline{A} \mid \overline{B}) = \frac{45}{75} = 0.60
\end{align*}
\]

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 \( A_i \) and \( B_i \) are independent if and only

\[
\begin{align*}
(i) \quad P(A_i \mid B_1) &= P(A_i \mid B_2) = P(A_i) \quad \text{and} \\
(ii) \quad P(A_2 \mid B_1) &= P(A_2 \mid B_2) = P(A_2)
\end{align*}
\]

The equation (i) of (1.5) says that neither the occurrence of \( B_1 \) nor \( B_2 \) affects the
occurrence of \( A_i \). 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\( \times \)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:

\[
P(A_1 \cap B_1) = P(A_2 \cap B_2) \text{ if and only if } P(A_1) = P(A_2), \ P(B_1) = P(B_2).
\]

This means that the \( 2 \times 2 \) 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:

\[
\begin{align*}
(a) \quad P(A_1 \mid B_1) &< P(A_1) < P(A_1 \mid B_2) \text{ iff } N \mid < 0 \\
(b) \quad P(A_1 \mid B_1) &= P(A_1) = P(A_1 \mid B_2) \text{ iff } N \mid = 0 \\
(c) \quad P(A_1 \mid B_1) &> P(A_1) > P(A_1 \mid B_2) \text{ iff } N \mid > 0
\end{align*}
\]

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

\[
\frac{n_{11}}{n_{11} + n_{21}} < \frac{n_{11} + n_{12}}{n} \quad \text{and} \quad \frac{n_{11} + n_{12}}{n} < \frac{n_{12}}{n_{12} + n_{22}}.
\]

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. \(|N| < 0\). \hspace{1cm} (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

\[
\frac{n_{11}}{n_{11} + n_{21}} < \frac{n_{11} + n_{12}}{n}, \quad \text{i.e.} \quad P(A_i \mid B_i) < P(A_i).
\]

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

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

(b) See Joarder (1998).

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

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

Part (b) implies that the events \(A_i\) and \(B_i\) 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) $n_{ij} = \frac{n_{i1}n_{j}}{n}$ 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 ($B$):

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$\overline{A}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>$\overline{B}$</td>
<td>36</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
</tr>
</tbody>
</table>

Here the following three probabilities

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

are not the same. Observe that $\mid N \mid < 0$ and $P(A \mid B) < P(A) < P(A \mid \overline{B})$. This means 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

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

are not the same. Observe that $\mid N \mid < 0$ and $P(B \mid A) < P(B) < P(B \mid \overline{A})$. This means that computers without modem boards are more likely to have printer boards than computers with modems. 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 ($B$):

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$\overline{A}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>$\overline{B}$</td>
<td>28</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60</td>
</tr>
</tbody>
</table>
Here the following three probabilities

\[ P(A \mid B) = \frac{12}{20} = 0.60, \quad P(A) = \frac{40}{100} = 0.40, \quad P(A \mid B^c) = \frac{28}{80} = 0.35 \]

are not the same. Observe that \( N \mid > 0 \) and \( P(A \mid B) > P(A) > P(A \mid B^c) \). This means 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

\[ P(B \mid A) = \frac{12}{40} = 0.30, \quad P(B) = \frac{20}{100} = 0.20, \quad P(B \mid A^c) = \frac{8}{60} = 0.13, \]

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

To explain (b) of Theorem 2.1, consider Example 1.1 that provides the following the breakdown of a computer having a modem board (A) or a circuit board (B):

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>( \bar{A} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>( \bar{B} )</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>40</td>
<td>60</td>
<td>100</td>
</tr>
</tbody>
</table>

Here the following three probabilities

\[ P(A \mid B) = \frac{10}{25} = 0.40, \quad P(A) = \frac{40}{100} = 0.40, \quad P(A \mid B^c) = \frac{30}{75} = 0.40 \]

are the same. Observe that \( N \mid = 0 \) and \( P(A \mid B) = P(A) = P(A \mid B^c) \). The same is true for the following three probabilities:

\[ P(B \mid A) = \frac{10}{40} = 0.25, \quad P(B) = \frac{2}{100} = 0.25, \quad P(B \mid A^c) = \frac{15}{60} = 0.25, \]

Observe that \( N \mid = 0 \) and \( P(B \mid A) = P(B) = P(B \mid A^c) \). 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 \( 2 \times 2 \) contingency table with categories of interest.
Acknowledgements

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

References
