A Counterexample in Probability

Chris Calabro
July 9, 2004

Let $A$ and $B_1, \ldots, B_n$ be events in some probability space and consider the following (invalid) inequality

$$Pr(A \mid \bigcup_i B_i) \geq \min\{Pr(A \mid B_i)\}. \quad (1)$$

1 Counterexamples

(1) is appealing because it seems to be saying, “Knowledge that one of the $B_i$ occurs cannot make $A$ less likely than knowledge that $B_j$ occurs, when $j$ is chosen so as to make $A$ as unlikely as possible.”

For example, let $n = 2$ and suppose

- $B_1$ is the event that a tornado occurs,
- $B_2$ is the event that a flood occurs,
- $A$ is the event that your house gets destroyed.

If someone allowed you to choose between knowledge of $B_1$, knowledge of $B_2$, and knowledge of $B_1 \cap B_2$, you might think that necessarily either knowledge of $B_1$ or knowledge of $B_2$ would make $A$ less likely, and so one of those options would always be correct.

But, on the contrary, it could happen that $B_1 \cup B_2$ is the best option. For suppose that

- $Pr(B_1) = Pr(B_2) = \frac{2}{3}$,
- $Pr(B_1 \cap B_2) = \frac{1}{3}$,
- $A = B_1 \cap B_2$.

so that $Pr(B_1 \cup B_2) = 1$. Then

$$Pr(A \mid B_1 \cup B_2) = Pr(A) = \frac{1}{3} < \frac{1}{2} = Pr(A \mid B_i)$$

for $i = 1, 2$. And so in this case, (1) is false. If these were your probabilities, then “there was a tornado or a flood” would be much better news than either of “there was a tornado” or “there was a flood”. 
Now consider an extreme example. Suppose $n$ is large, $A = \bigcap_i B_i$, and each $B_i$ is the union of 2 parts, $A$ and $C_i$, where the measure of $C_i$ is small compared to that of $A$, the $C_i$ are pairwise disjoint, and the measure of $A$ is small compared to 1.

Geometrically, we can imagine this as $A$ is a sphere, the $C_i$ are thin but very numerous spikes coming out of the sphere (so that together they have much larger volume than $A$ does), and $B_i = A \cup C_i$. Then the lhs of (1) is nearly 0, while the rhs is nearly 1.

To make this seem more relevant, let us give real-life meanings to the variables: $n$ is the number of people playing the lottery, $C_i$ is the event that person $i$ wins, and $A$ is the event that no one wins so that the state keeps the money. Then $B_i$ is the event that either $i$ wins or no one wins. The state tries to set things up so that usually someone wins, (or at least let us pretend that this is the case!) so that $Pr(A \mid \bigcup_i B_i)$ is small, say $\frac{1}{100}$. But for each fixed person $i$, given that (either $i$ wins or no one wins), the probability that $i$ wins is very small.

2 Sufficient hypothesis

If we assume that the $B_i$ are pairwise disjoint, then (1) holds. This follows easily by induction on $n$ provided we can show the $n = 2$ case. The goal is to show that

$$Pr(A \mid B_1 \cup B_2) \geq Pr(A \mid B_1)$$

or $Pr(A \mid B_1 \cup B_2) \geq Pr(A \mid B_2)$,

which is equivalent to

$$\frac{Pr(A \cap B_1) + Pr(A \cap B_2)}{Pr(B_1) + Pr(B_2)} \geq \frac{Pr(A \cap B_1)}{Pr(B_1)}$$

or

$$\frac{Pr(A \cap B_1) + Pr(A \cap B_2)}{Pr(B_1) + Pr(B_2)} \geq \frac{Pr(A \cap B_2)}{Pr(B_2)},$$

making use of pairwise disjointness and assuming (wlog?) $Pr(B_1), Pr(B_2) > 0$.

If we let $a = Pr(A \cap B_1), b = Pr(A \cap B_2), c = Pr(B_1), d = Pr(B_2)$, then it is sufficient to show, given $a, b \geq 0, c, d > 0$, that

$$\frac{a + b}{c + d} \geq \frac{a}{c}$$

or

$$\frac{a + b}{c + d} \geq \frac{b}{d},$$

which is equivalent to

$$(a + b)c \geq a(c + d)$$

or $$(a + b)d \geq b(c + d),$$

2
which is equivalent to

\[ bc \geq ad \]

or \( ad \geq bc \),

which, evidently, is true.

3 Another perspective

The same problem can appear in another guise: suppose we know that \( \forall i \in [n] \)

\[ \Pr(A \mid B_i) \geq \Pr(A), \]

can we conclude that \( \Pr(A \mid \bigcup_i B_i) \geq \Pr(A) \)? It seems reasonable. After all, if each \( B_i \) is good news, then surely the 'or' of the \( B_i \) must be good news as well.

But this is false. For consider the following counterexample. Let the probability space be the tetris piece shaped like a 'T' (3 blocks on the bottom row and 1 block in the middle of the top row). Let \( A \) be the 2 \( \times \) 1 vertical domino, \( B_1 \) be the 1 \( \times \) 2 left horizontal domino, and \( B_2 \) be the 1 \( \times \) 2 right horizontal domino.

More formally,

- The sample space is \( a \cup b \cup c \cup d \) where \( a, b, c, d \) are pairwise disjoint and each has probability \( \frac{1}{4} \).
- \( A = a \cup c \)
- \( B_1 = b \cup c \)
- \( B_2 = c \cup d \).

Then \( \Pr(A) = \frac{1}{2} \), \( \Pr(A \mid B_i) = \frac{1}{2} \), and yet \( \Pr(A \mid B_1 \cup B_2) = \frac{1}{3} \).

This can be phrased as the problem of the king’s sibling: the king has 1 sibling; what is the probability that the sibling is male? It is \( \frac{1}{3} \), which may sound strange at first. But let the sample space be \( \{(b, b), (b, g), (g, b), (g, g)\} \) to represent the genders of a random pair of children. \( A \) be the event that the children have the same gender, \( B_1 \) be the event that the first born is a boy, and \( B_2 \) be the event that the second born is a boy. Then \( \Pr(A) = \frac{1}{2}, \Pr(A \mid B_i) = \frac{1}{2} \), and \( \Pr(A \mid B_1 \cup B_2) = \frac{1}{3} \).