Dependence of events, revisited

PETER BRAZA

Introduction

Independence is a key concept in probability. Conceptually, we think of two events as being independent if the outcome of one event doesn't affect the outcome of the other and vice versa. Mathematically, we say that events A and B are independent if the probability that both occur is the product of the probabilities that each occurs. More precisely, $P(A \cap B) = P(A)(P(B)$ in which $P()$ denotes the probability of the given event. Alternatively, we say that A and B are independent if the conditional probability that A occurs given that B has occurred, $P(A \mid B)$, satisfies $P(A \mid B) = P(A)$. That is, whether or not B occurs does not affect whether or not A occurs.

For example, if a coin is flipped twice and A is the event that a head occurs on the first toss and B is the event that a head occurs on the second toss, then A and B do not depend on each other – they are independent events. In this case, $P(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A) \times P(B)$.

For a slight twist, what if A is the event that at least one head appears on the two tosses and B is the event of getting two heads? Conceptually, the two events depend on each other since the only way B can occur is if A occurred (but note, not vice versa). Mathematically, $P(A \cap B) = \frac{1}{4} \neq \frac{3}{4} \times \frac{1}{4} = P(A) \times P(B)$.

Textbooks and articles on probability are replete with examples of independent events and random variables. The alternative to independence is typically just given as 'not independent' or 'dependent'. For nuance, why not include a mathematical measure of how dependent one event is on another that captures the spectrum of possibilities from 'a tiny amount' to 'completely dependent'? A few articles and books address levels of dependence (see [1] to [5]). However, the methods in [1] and [2] are highly technical and require advanced mathematical tools. The authors in [3, 4, 5] discuss the independence/dependence of events by connecting events to random variables via the indicator function and then using the Pearson correlation coefficient for random variables (hereinafter referred to as the correlation coefficient) as a surrogate measure for the dependence of events. This approach makes sense but would benefit from an alternative formulation derived directly for events.

How could one arrive at a comparable definition of the dependence of events without resorting to the structure that goes into the correlation coefficient? The purpose of this paper is fourfold:

(1) to answer this question by going through an elemental process of mathematical discovery that naturally leads to a measure of the dependence of one event on another;

(2) to illustrate how well the measure comports with our intuitive sense of how much events depend on each other;

(3) convince teachers to supplement their probability lessons with material on dependence to deepen student understanding of events;

https://doi.org/10.1017/mag.2024.64 Published online by Cambridge University Press

(4) to invite the reader to create their own examples of events with various levels of dependence – and to have fun doing so!

Coming up with a measure of dependence for events

A starting approach to come up with a measure for the dependence of two events A and B is to see how much the events deviate from being independent. Mathematically, this could mean considering the size of the differences $P(A \cap B) - P(A)P(B)$ or $P(A \mid B) - P(A)$. For independent events, those differences are zero. So far, so good, but these differences by themselves have drawbacks to being good measures of dependence. As an example, the first difference will be small when A and B have small probabilities, automatically implying that A and B are somehow nearly independent. Yet, if $B = A$, then it only makes sense to think of A and B as being completely dependent. This suggests that we need some normalising factor in the differences above. Also, a problem with the latter difference above is that $P(B \mid A) - P(B)$ is equally reasonable, but $P(A \mid B) - P(A)$ is generally not the same as $P(B \mid A) - P(B)$.

Normalising factors are commonly used in mathematics and other fields. As examples, the decibel scale for sound and the Richter scale for earthquakes use reference intensities. Probability density functions employ normalising constants to ensure that the overall probability of a sample space is 1. With this in mind, we include a normalising factor and account for the asymmetry of $P(A \mid B)$ and $P(B \mid A)$ to suggest the following as *possible* measures for the dependence of two events:

$$\frac{P(A \mid B) - P(A)}{P(A)} \text{ or } \frac{P(B \mid A) - P(B)}{P(B)}, \tag{1}$$

which are equivalent to

$$\frac{P(A \cap B) - P(A)P(B)}{P(A)P(B)}. \tag{2}$$

Of course, we need to require $P(A) \neq 0$ and $P(B) \neq 0$, although we could account for these cases by way of limits. Expressions (1) and (2) can be negative or positive. This is a desirable feature since an eventual goal for a measure of dependence would be for it to lie between the normalised values of -1 and 1 to correspond to the correlation coefficient for random variables ([3 to 7]). However, this goal isn't satisfied because of the following shortcoming with the expressions: If say, $B \subseteq A$, then they become $\dfrac{1 - P(A)}{P(A)}$, which gets arbitrarily large as $P(A)$ goes to zero.

To overcome this mathematical obstacle, we incorporate the idea that A and B are independent when the probability of A does not depend on whether B occurs or does *not* occur. In mathematical terms, if A and B are independent events, then so are A and B^C, in which B^C denotes the complement of B ([5 to 7]). Likewise, A^C and B are independent. Basically, making use of the events A^C and B^C uses all the information in the sample space, so their probabilities are important to include when measuring

https://doi.org/10.1017/mag.2024.64 Published online by Cambridge University Press

dependence. As such, we consider the following expression of dependence of two events A and B:

$$\frac{P(A \cap B) - P(A)P(B)}{P(A)P(B)} \times \frac{P(A^C \cap B^C) - P(A^C)P(B^C)}{P(A^C)P(B^C)}. \tag{3}$$

This expression can be visually simplified by the following substitutions: $P(A) = a$, $P(B) = b$ and $P(A \cap B) = c$ so that $P(A^C) = 1 - a$, $P(B^C) = 1 - b$ and $P(A^C \cap B^C) = 1 + c - a - b$. The last equality is true because $P(A^C \cap B^C) = P((A \cup B)^C) = 1 - (a + b - c)$.

Putting all this together and simplifying gives the following tidy expression for a coefficient of dependence, D, for two events A and B:

$$D = \frac{c - ab}{\sqrt{ab(1 - a)(1 - b)}}. \tag{4}$$

When taking the square root of (3), we used the same numerator as in (2) for the value of D.

Note that when two events are independent, then $D = 0$ since $P(A \cap B) = P(A)P(B)$. As it importantly turns out, the expression for D equates to the correlation coefficient for random variables defined by indicator functions for the events A and B (see [3, 4, 5]). We insert a caution here: note that when considering more general cases than two simple events, the correlation coefficient can take the value zero for dependent variables. However, other measures of dependence can be used to avoid this potential drawback.

Establishing $-1 \leqslant D \leqslant 1$

The result $-1 \leqslant D \leqslant 1$ follows from the fact that the correlation coefficient is between -1 and 1; this is usually proved using the Cauchy-Schwarz inequality. However, in the spirit of self-contained analysis we show how to get this result independently.

To begin, we can assume that $b \leqslant a$, without loss of generality. Since $c = P(A \cap B) \leqslant b$ and the fact that the probability of any event is at least 0 and at most 1, we get the extended inequality, $0 \leqslant c \leqslant b \leqslant a \leqslant 1$.

Observe that the minimum value of D occurs when $c = 0$, and the maximum occurs when $c = b$ (for fixed a and b). When A and B are mutually exclusive, then $c = 0$, so that $D^2 = \dfrac{ab}{(1 - a)(1 - b)}$. Also, $c = 0$ gives us the additional constraint, $a + b \leqslant 1$, since the sum of the probabilities of A and B cannot exceed 1. Rewriting D^2 as

$$D^2 = 1 - \frac{1 - (a + b)}{1 - (a + b) + ab},$$

we see that the second term is non-negative and less than or equal to 1, which implies that $|D| \leqslant 1$. The case $c = b$ occurs when $B \subseteq A$, and it results in $D^2 = \dfrac{1 - a}{a}\dfrac{b}{1 - b}$. This is a function of two variables defined over

https://doi.org/10.1017/mag.2024.64 Published online by Cambridge University Press

the triangular region $0 \leqslant b \leqslant a \leqslant 1$. As it turns out, standard calculus techniques show that the maximum value of D^2 in this region is 1, which occurs on the boundary component, $a = b$, of the region. In sum, the maximum value of $|D|$ is 1 in both cases; therefore D satisfies $-1 \leqslant D \leqslant 1$.

This inequality brings up the notion of positive and negative dependence. Positive dependence ($D > 0$) occurs when $P(A \cap B) > P(A)P(B)$. Likewise, events A and B are negatively dependent when $P(A \cap B) < P(A)P(B)$.

Examples

The test of whether D is a good measure of the dependence of events is how well it agrees with our intuitive sense of how much one event depends on another. A variety of examples will illustrate this.

As a first example, we consider the case when $A = B$. Obviously, they are completely dependent, and it is the case that $D = 1$ (assuming $a = b > 0$). However, if $B = A^C$, we get $D = -1$ which indicates a complete negative dependence. This makes sense because A and B comprise the whole sample space, yet they have no elements in common.

In the second example in the introduction in which a coin is tossed twice and A is the event that at least one head appears while B is the event that two heads appear, we get that A is the set $A = \{HH, HT, TH\}$ and $B = \{HH\}$. (The first/second element of each entry is the outcome of the first/second coin toss). As such, $a = \frac{3}{4}$, $b = \frac{1}{4}$ and $c = \frac{1}{4}$, yielding a coefficient of dependence of $D = \frac{1}{3}$. As a related yet somewhat contrasting example, suppose A is the event that a head appears on the first toss and B is the same as above; in this case $A = \{HH, HT\}$ and $B = \{HH\}$. Conceptually A and B are more dependent on each other in this example than in the prior example, and this is borne out in the coefficient of dependence: $D = \frac{\sqrt{3}}{3} \approx 0.577$. Note that these two examples are special cases of the general case when $B \subseteq A$, as discussed above.

As a classical example, suppose one picks two marbles from a jar that contains g green marbles and r red marbles. Suppose A is the event that the first marble is red, and B is the event that the second marble is green. If the first marble is replaced before picking the second, then A and B are independent so that $D = 0$. However, let us consider the case in which the first marble is not replaced. If we let $n = g + r$, then $P(A) = \frac{r}{n}$ and $P(B) = \frac{g}{n}$, in which the latter probability is obtained from Bayes' formula: $P(B) = P(B \mid A)P(A) + P(B \mid A^C)P(A^C) = \frac{g}{n-1}\frac{r}{n} + \frac{g-1}{n-1}\frac{g}{n} = \frac{g}{n}$. Furthermore, the probability of picking a red marble first and then a green is $P(A \cap B) = \frac{r}{n}\frac{g}{n-1}$; this does not equal $P(A)P(B) = \frac{r}{n}\frac{g}{n}$, so A and B are not independent. The coefficient of dependence is calculated to be $D = \frac{1}{n-1}$. The interpretation is that as the number of marbles (n) in the jar increases, the dependence of the two events decreases and goes to zero as the number of the marbles gets very large. Basically, not replacing the first marble becomes irrelevant from a dependence standpoint when the total number of marbles is large.

https://doi.org/10.1017/mag.2024.64 Published online by Cambridge University Press

As a further check on the reasonableness of D as a measure of the dependence of events, we consider the case when the events A and B are mutually exclusive; i.e. $A \cap B = \varnothing$ This implies $c = 0$ in (4), so that $D = -\sqrt{\frac{ab}{(1-a)(1-b)}}$. Firstly, D is negative (assuming $a, b > 0$), signifying that A and B are negatively dependent. This agrees with our intuition since A and B have no elements in common. To further marshal our thoughts, we illustrate two cases. If a and b are both small, then D is also small, which implies A and B are nearly independent. This is expected since A and B do not comprise much of the sample space. To get values of D at the other end of the D spectrum (D close to -1), we consider events A and B such that $A \cup B$ comprises all of (or most of) the sample space; this implies $P(A \cup B) \approx 1$. Since $P(A \cup B) = P(A) + P(B)$ when A and B are mutually exclusive, we have that $a + b \approx 1$. If $a + b$ is 1 or close to 1, then D is -1 or close to -1 as well. These values for D makes sense since $A \cup B$ fills the sample space, yet A and B have no elements in common.

Concluding Comments

The measure of dependence, D, gives a scrutable way of characterising two events by giving them a numerical place along the spectrum from negative dependence to independent to positive dependence. It is easy to calculate D in all sorts of probability problems, and it can be both informative and fun to see how D agrees with our sense of how dependent two events are. At a fundamental level, knowing the value of D is much more satisfying than merely saying independent or dependent.

References

1. A. Rényi, On measures of dependence, *Acta Math. Acad. Sci. Hungary.* **10,** pp. 441- 451 (1959).
2. Hui Li, A true measure of dependence, *Munich Personal RePEc Archive* (2016), available at https://mpra.ub.uni-muenchen.de/69735/
3. Stepniak Czeslaw, A note on correlation coefficient between random events, *Discussiones Mathematicae, Probability and Statistics*, **35** pp 57-60 (2015).
4. L.T. Kubik, *Probability, A textbook for teaching mathematics studies* (in Polish), Polish Scientific Publishers (1980).
5. Jim Pitman, *Probability*, Springer-Verlag (1993).
6. Sheldon Ross, *A first course in probability*, Collier Macmillan (1976).
7. Morris H. DeGroot, *Probability and statistics*, Addison-Wesley (1975).

PETER BRAZA
University of Colorado,
Colorado Springs,
1420 Austin Bluffs Road,
Colorado Springs, CO 80918 USA
e-mail: *pbraza@uccs.edu*