



Why

We want to talk about how knowledge of one aspect of an outcome can give us knowledge about another aspect.

Dependent events

Suppose Ω is a finite set of outcomes with event probability function P . Two uncertain events A and B with $P(A), P(B) > 0$ are *dependent* under P if

$$P(A \mid B) \neq P(A) \quad \text{or} \quad P(B \mid A) \neq P(B)$$

In other words, if conditioning on one will change the probability of the other. We can rewrite the condition as,

$$\frac{P(A \cap B)}{P(B)} \neq P(A) \quad \text{or} \quad \frac{P(A \cap B)}{P(A)} \neq P(B)$$

We see that in either case, $P(A \cap B) \neq P(A)P(B)$.

Definition

Two events A and B are *independent* under P if

$$P(A \cap B) = P(A)P(B).$$

In other words, they are independent if the probability of their intersection is the product of their respective probabilities. This definition clearly shows that independence (and dependence) is a *symmetric* relation on the set of events. Moreover, the expression $P(A \cap B) = P(A)P(B)$ is well-defined even when $P(A)$ or $P(B)$ is 0.

As we have seen, in the case that $P(B) \neq 0$, $P(A \cap B) = P(A)P(B)$ is equivalent to $P(A \mid B) = P(A)$. Roughly speaking, we interpret this second expression as encoding the fact that the occurrence of event B does not change the probability—intuitively, the “credibility”—of the event A .

Examples

Two coin tosses. As usual, model flipping a coin twice with the sample space $\Omega = \{0, 1\}^2$. Define $A = \{(1, 0), (1, 1)\}$, the event that the first toss is heads, and $B = \{(0, 1), (1, 1)\}$, the event that the second toss turns up head. Then $A \cap B$ is $\{(1, 1)\}$, the event both tosses turn up heads. Suppose we put a distribution on Ω as usual with $p(\omega) = 1/4$ for all $\omega \in \Omega$. Then $P(A) = 1/2$, $P(B) = 1/2$, and $P(A \cap B) = 1/4$. Thus:

$$P(A \cap B) = P(A)P(B)$$

and so the events are independent events under the distribution.

Three tosses. As usual, model flipping a coin three times with the sample space $\Omega = \{0, 1\}^3$ and define $p : \Omega \rightarrow [0, 1]$ by $p(\omega) = 1/8$ for all $\omega \in \Omega$. Let A be the event $\{(1, 1, 0), (1, 1, 1)\}$, the event that the first two tosses turn up heads, and $B = \{(1, 0, 0), (0, 0, 0)\}$, the event the second two tosses are tails. Then $P(A) = P(B) = 2/8 = 1/4$, but $A \cap B = \emptyset$. So $P(A)P(B) = 1/16 \neq 0P(A \cap B)$. These are dependent events under the model.

Rolling two dice. As usual, model rolling two dice with the sample space $\Omega = \{(\omega_1, \omega_2) \mid \omega_i \in \{1, \dots, 6\}\}$. Define a distribution $p : \Omega \rightarrow \mathbf{R}$ by $p(\omega) = 1/36$. Two events are $A = \{\omega \in \Omega \mid \omega_1 + \omega_2 > 5\}$, “the sum is greater than 5”, and $B = \{\omega \in \Omega \mid \omega_1 > 3\}$, “the number of pips on the first die is greater than 3”. Then $P(A) = 26/36$. Also, $P(A \mid B) = 17/16$. So, these events are dependent. Roughly speaking, we say that knowing B tells us something about A . In this case, we say that it “makes A more probable.” In the language used to describe the events, we say that knowledge that the number of pips on the first die is greater than three makes it more probable that the sum of the number of pips on each die is greater than 5.

Basic implications

Suppose P is a probability measure on a finite sample space Ω . It happens that A and B 's independence under P is sufficient for the independence of A^c and B , A^c and B^c and A and B^c . Here A^c and B^c denote the relative complement of A and B in Ω , respectively.

There are a few ways to see these—here's one

$$\begin{aligned} P(A^c \cap B) &= P(B) - P(A \cap B) \\ &= P(B) - P(A)P(B) \\ &= (1 - P(A))P(B) = P(A^c)P(B). \end{aligned}$$

The first equality here holds since $P(B) = P(A \cap B) + P(A^c \cap B)$.

Here are alternative routes, using the more explicit notation of relative complements. To this, we note that since $P(\cdot \mid B)$ is a probability measure, and the events A and $C_\Omega(A)$ partition Ω , we have

$$P(A \mid B) + P(C_\Omega(A) \mid B) = 1.$$

From which we deduce $P(C_\Omega(A) \mid B) = 1 - P(A \mid B) = P(C_\Omega(A))$. Which is equivalent to $P(C_\Omega(A) \cap B) = P(C_\Omega(A))P(B)$. In other words, B and $C_\Omega(A)$ are independent events. Similarly, A and $C_\Omega(B)$ are independent events. Since $(\Omega - A) \cap (\Omega - B) = \Omega - (A \cup B)$, we have

$$P(A \cup B) + P(C_\Omega(A) \cap C_\Omega(B)) = 1.$$

Since $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, we obtain

$$P(C_\Omega(A) \cap C_\Omega(B)) = 1 - P(A) - P(B) + P(A \cap B) = P(A)P(B).$$

We can express the right hand side as $(1 - P(A))(1 - P(B))$ or $P(C_\Omega(A))P(C_\Omega(B))$. In other words, $C_\Omega(A)$ and $C_\Omega(B)$ are independent.

