Conditional Probability

It is commonly agreed that conditional probability is one of the most fundamental concepts in the theory, because, firstly, it captures the idea of partial information about an event E in the form of another event F.

ex. 1 Let $S = \{(i,j) \mid i,j = 1,2,3,4,5,6\}$ be the sample space of outcomes of rolling 2 dice, and let

$$E = \{(i,j) \mid i+j = 8,\ i,j = 1,\dots,6\}$$
$$= \{(2,6),\ (3,5),\ (4,4),\ (5,3),\ 6,2)\}$$

$$F = \{(i,j) \mid i = 3,\ j = 1,\dots,6\}.$$

Q: If we have the partial information that F has occurred, what is $P(E)$?

A: Since $F = \{(i,j) \mid i=3,\ j=1,\dots,6\}$

$$= \{(3,1),\ (3,2),\ (3,3),\ (3,4),\ (3,5),\ (3,6)\}$$

has 6 elements, only one of which lies in E, namely $(3,5)$, and since each one of the 36 outcomes (i,j) has ~~prob~~ equal probability of $\frac{1}{36}$,

$$P(E|F) = \frac{1/36}{6/36}$$

$1/36 = P(E \cap F) = P(\{(3,5)\})$

$6/36 = P(F)$

$$= \boxed{\frac{1}{6}}$$

Alternatively, since $E \cap F = \{(3,5)\}$, and $|F| = 6$,

$$P(E|F) = \boxed{\frac{1}{6}}$$

$\boxtimes$

The <u>conditional probability</u> that E will occur <u>given that F has occurred</u> is

$$\boxed{P(E|F) \stackrel{\text{def}}{=} \frac{P(E\cap F)}{P(F)}}$$

provided $P(F) > 0$ (if $P(F) = 0$, then $P(E|F)$ is undefined).

<u>Prop. 1</u> (<u>multiplication rule</u>)

$$\boxed{P(E\cap F) = P(F)P(E|F)}$$

More generally, for events $E_1, E_2, \ldots, E_n \in \mathcal{A}$,

$$\boxed{\begin{aligned} &P(E_1\cap E_2\cap \cdots \cap E_n) \\ &\quad = P(E_1)P(E_2|E_1)P(E_3|E_1\cap E_2) \\ &\qquad \cdots P(E_n|E_1\cap E_2\cap\cdots\cap E_{n-1}) \end{aligned}}$$

pf: Since $P(E|F) = \dfrac{P(E \cap F)}{P(F)}$, we immediately have

$$P(E \cap F) = P(F)\,P(E|F)$$

Similarly, since

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

$$P(E_3|E_1 \cap E_2) = \frac{P(E_1 \cap E_2 \cap E_3)}{P(E_1 \cap E_2)}$$

$$\vdots$$

$$P(E_n|E_1 \cap \cdots \cap E_{n-1}) = \frac{P(E_1 \cap \cdots E_{n-1} \cap E_n)}{P(E_1 \cap \cdots \cap E_{n-1})}$$

we see that the RHS simplifies to the LHS:

$$P(E_1)P(E_2|E_1) \cdots P(E_n|E_1 \cap \cdots \cap E_{n-1})$$

$$= \cancel{P(E_1)} \cdot \frac{\cancel{P(E_1 \cap E_2)}}{\cancel{P(E_1)}} \cdot \frac{\cancel{P(E_1 \cap E_2 \cap E_3)}}{\cancel{P(E_1 \cap E_2)}} \cdots \frac{\cancel{P(E_1 \cap \cdots \cap E_n)}}{\cancel{P(E_1 \cap \cdots \cap E_{n-1})}}$$

does not cancel

$$= P(E_1 \cap \cdots \cap E_n) \qquad \boxtimes$$

Bayes' Formula

Prop. 2 Let $E \in \mathcal{A}$ be any event. Given any other event $F \in \mathcal{A}$,

$$\boxed{P(E) = \underbrace{P(E|F)P(F) + P(E|F^c)(1-P(F))}_{\text{weighted average of conditional probs.}}}$$

weights

More generally, if $F_1, \dots, F_n \in \mathcal{A}$ are mutually disjoint events, & $S = F_1 \sqcup F_2 \sqcup \cdots \sqcup F_n$, then

$$\boxed{P(E) = \overbrace{\sum_{i=1}^{n} P(E|F_i)P(F_i)}^{\text{weighted average}}}$$

(law of total probability)

weights

pf: $S = F \sqcup F^c \Rightarrow E = E \cap S$

$$= E \cap (F \sqcup F^c)$$

$$= (E \cap F) \sqcup (E \cap F^c)$$

$$\Rightarrow P(E) = P((E \cap F) \sqcup (E \cap F^c))$$

$$= P(E \cap F) + P(E \cap F^c)$$

$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

Similarly,

$$S = \bigsqcup_{i=1}^{n} F_i \implies E = E \cap S$$
$$= E \cap \left(\bigsqcup_{i=1}^{n} F_i \right)$$
$$= \bigsqcup_{i=1}^{n} (E \cap F_i)$$

$$\implies P(E) = P\left(\bigsqcup_{i=1}^{n} E \cap F_i \right)$$
$$= \sum_{i=1}^{n} P(E \cap F_i)$$
$$= \sum_{i=1}^{n} P(E \mid F_i) P(F_i)$$

<u>QED</u>

<u>Prop. 3</u> Let $E \in \mathcal{A}$ & let $F_1, \ldots, F_n \in \mathcal{A}$ be mutually exclusive & exhaustive $\left(S = \bigsqcup_{i=1}^{n} F_i\right)$.

Then,

$$\boxed{\begin{aligned} P(E \mid F_j) &= \frac{P(E \cap F_j)}{P(E)} \\ &= \frac{P(E \mid F_j) P(F_j)}{\sum_{i=1}^{n} P(E \mid F_i) P(F_i)} \end{aligned}}$$

Independent Events

Let us first recall the definition of a conditional probability of an event E given information / evidence that another event F has occurred,

$$\boxed{P(E|F) = \frac{P(E \cap F)}{P(F)}} \qquad (1)$$

from which we get the multiplication rule,

$$\boxed{P(E \cap F) = P(E|F)P(F)} \qquad (2)$$

If F is pertinent to E, esp. to $P(E)$, we generally do not have $P(E) = P(E|F)$, since F affects $P(E|F)$. If, however, the occurrence of F does not affect $P(E)$, we say that E and F are *independent events*. In this case the multiplication rule becomes

$$\boxed{P(E \cap F) = P(E)P(F)} \qquad (3)$$

If (3) holds, E and F are called <u>independent</u>, else (if only (2), but not (3)) they are called <u>dependent</u>.

<u>Three events</u> E, F and G are called <u>independent</u> if

- $P(E \cap F \cap G) = P(E)P(F)P(G)$ } all 3 at once
- $P(E \cap F) = P(E)P(F)$
- $P(E \cap G) = P(E)P(G)$
- $P(F \cap G) = P(F)P(G)$

} pairwise indep. (4)

and in general events $E_1, E_2, \ldots, E_n$ are independent if

$$\boxed{\begin{aligned} &P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) \\ &\quad = P(E_{i_1})P(E_{i_2})\cdots P(E_{i_k}) \end{aligned}} \qquad (5)$$

for all $k = 2, \ldots, n$, where $i_1 < i_2 < \cdots < i_k$.

We can even call an infinite sequence of events independent if every finite subset of events is independent in the sense of (5).

In the case of a sequence of "subexperiments" which are part of one larger experiment (e.g. tossing a coin repeatedly), which we consider events $E_1, E_2, \ldots,$ we say the subexperiments E_i are *independent* in the above sense of (5), *if the ith subexperiment completely determines* $P(E_i)$. In this case we call the subexperiments *trials.*