ENEE 324 Engineering Probability Lecture 3

Conditioning

In a chance experiment $\mathcal{E}$, occurrence of event A can be influenced by that of event B. For instance, event $A = flooding$, always occurs following event $B = dam - burst$; event $A = flooding$ may have only a small likelihood of occurrence following event $C = light\, shower$.

Interdependence of events influences probabilities. Probabilities computed after obtaining data on one event can be different (higher or lower than) the probabilities computed before such data was available.

In weather forecasting, forecasts for a *Wednesday* made on *Tuesday, 8:00 AM* and *Tuesday, 8:00 PM* differ - additional observations are available during the intervening 12 hour period.

In a medical setting, the presence of a disease in a patient would increase the probability of certain symptoms in the patient. Some symptoms may be present even in the absence of disease. For example, certain symptoms are shared by allergies and by the common cold. In medical diagnosis, a doctor seeks to determine the probability of a certain disease being present given that certain symptoms are observed. This probability may be higher than when the symptoms are not observed. Thus one could say that the observation of a symptom influences the likelihood of a diagnosis of a disease. But this does not imply a causal relationship. Symptoms do not cause diseases!

Data *conditions* probabilities. In fact, practically all probabilities are *conditional probabilities*. We now give a formal definition.

Definition: Let $\mathcal{E}$ be a chance experiment with associated sample space Ω and Boolean algebra $\mathcal{A}$ of interesting events (*i.e.*, subsets of Ω). Given events $A, B \in \mathcal{A}$, the conditional probability of A given B denoted $P(A \mid B)$ is defined to be

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

provided $\quad P(B) \neq 0 \quad \square$

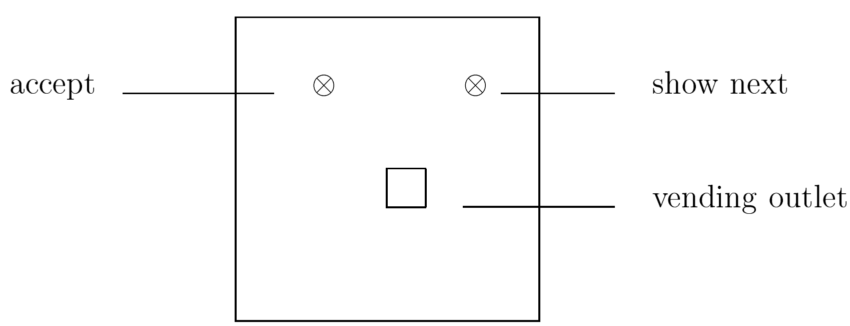
Note: If $P(B) = 0$, then $P(A \mid B)$ is *undefined*.

What is the justification for making such a definition? A bit of counting helps. Suppose experiment $\mathcal{E}$ has n equally probable/likely elementary events. Suppose n_A is the number of such elementary events favorable to the occurrence of event A. Suppose n_B is the number of elementary events favorable to the occurrence of event B. Then $P(B) = n_B/n$. If B actually occurs then the outcomes have to be one of n_B possibilities. Now, for A to occur, one looks at a subset of these that favor A, and these are $n_{(A\cap B)}$ of these. So it makes sense to say,

$$\begin{aligned} P(A \mid B) &= \frac{n_{A\cap B}}{n_B} \\ &= \frac{n_{A\cap B}/n}{n_B/n} \\ &= \frac{P(A \cap B)}{P(B)} \end{aligned}$$

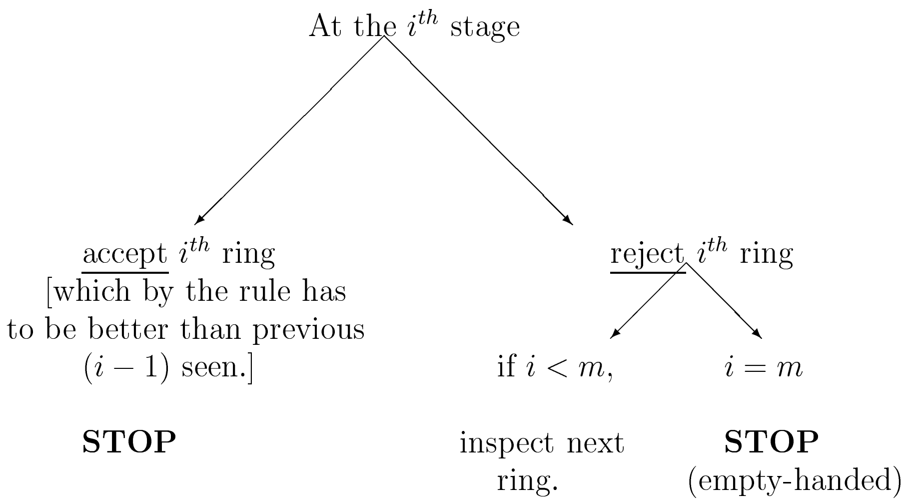
So our definition makes sense.

Example (optimal choice)



Consider a robot merchant that displays diamond rings one at a time to a player. There is a total of m rings. The order of presentation is random. The diamond rings are of differing quality. The player follows the rule: *Never accept a ring inferior to those previously rejected.* The player can press the

show next button or the **accept** button, until there are no rings remaining to be shown.



Question: Suppose the player selects the i^{th} ring. What is the probability of this being the best of all m rings? [This is a prototypical problem of deciding when to commit to a particular course of action or choice.]

Solution:

$$\begin{aligned} B \ :&= \ \text{event that the last of } i \text{ inspected} \\ &\quad\ \text{rings is the best of those inspected} \\ A \ :&= \ \text{event that the } i^{th} \text{ ring} \\ &\quad\ \text{is the best of all } m \text{ rings} \end{aligned}$$

We are interested in $P(A \mid B)$.
Clearly $A \subset B$. Hence $A \cap B = A$.
Thus,

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)}{P(B)} \end{aligned}$$

But

$$P(B) = \frac{(i-1)!}{i!} = \frac{1}{i}$$

Why? $(i-1)!$ is the number of permutations of i distinct things, leaving one, "the best ring," fixed in the i^{th} place.

$$P(A) = \frac{(m-1)!}{m!} = \frac{1}{m}$$

Why? $(m-1)!$ is the number of permutations of m distinct things, leave one, "the best ring," fixed in the i^{th} place. Thus,

$$P(A \mid B) = \frac{1/m}{1/i} = \frac{i}{m}$$

Late commitment is more likely to give you the best deal. □

Example: Toss 2 fair dice, producing the outcome (X, Y). Here, $X, Y \in \{1, 2, 3, 4, 5, 6\}$. Consider the events,

$$\begin{aligned} A &= \{(X, Y) \mid X + Y = 10\} \\ B &= \{(X, Y) \mid X > Y\} \end{aligned}$$

What is the probability $P(A \mid B)$?

$$\begin{aligned} B &= \{(2,1),\ (3,1),\ (3,2),\ (4,1),\ (4,2),\ (4,3), \\ &\quad (5,1,),\ (5,2),\ (5,3),\ (5,4),\ (6,1),\ (6,2), \\ &\quad (6,3),\ (6,4),\ (6,5)\} \\ n_B &= 15 \end{aligned}$$

Conditioning on B means one can *reduce* the sample space from the full set Ω of all 36 ordered pairs (X, Y) to the smaller subset B.
Within B, there is only one outcome $(6, 4)$ yielding $6 + 4 = 10$, favorable to A. So $n_{(A \cap B)} = 1$.

$$\begin{aligned} P(A \mid B) &= \frac{n_{A \cap B}}{n_B} \\ &= \frac{1}{15}. \end{aligned}$$

But $A = \{(6,4),\ (4,6),\ (5,5)\} \Rightarrow P(A) = \dfrac{3}{36} = \dfrac{1}{12}$

Also, $\dfrac{P(A \cap B)}{P(B)} = \dfrac{n_{A \cap B}/n}{n_B/n} = \dfrac{1/36}{15/36} = \dfrac{1}{15}$, as we expect.

Thus, $P(A)$ is different from $P(A \mid B)$.

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{1/36}{1/12} = \frac{1}{3}.$$

Properties of Conditional Probability

(1) $0 \le P(A \mid B) \le 1$.

Proof: $\emptyset \subset A \cap B \subset B$. Hence, $P(\emptyset) \le P(A \cap B) \le P(B)$.
It follows that $0 \le \dfrac{P(A \cap B)}{P(B)} \le 1 \quad \square$.

(2) $A \cap B = \emptyset$. Then $P(A \mid B) = 0$.

(3) $B \subset A$, then $P(A \mid B) = 1$.

Proof: $B \subset A \Rightarrow B \cap A = B$. Thus,

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(B)}{P(B)} \\ &= 1 \qquad \square \end{aligned}$$

(4) Given $A_1, A_2, \cdots A_k$ disjoint, and $A = \cup_{i=1}^k A_i$. Then, $P(A \mid B) = \sum_{i=1}^{k} P(A_i \mid B)$.

Proof: $A \cap B = \left(\cup_{i=1}^k A_i\right) \cap B = \cup_{i=1}^k (A_i \cap B)$.
Since A_i are disjoint, $A_i \cap B$ are also disjoint. Thus,

$$P(A \cap B) = P((\cup_{i=1}^k A_i) \cap B)$$

$$
\begin{aligned}
&= \sum_{i=1}^{k} P(A_i \cap B). \\
\text{Hence } P(A \mid B) &= \sum_{i=1}^{k} \frac{P(A_i \cap B)}{P(B)} \\
&= \sum_{i=1}^{k} P(A_i \mid B) \qquad \square
\end{aligned}
$$

We have shown that conditional probability of a disjoint union is the sum of the conditional probabilities. This demonstrates the parallel to the addition axiom for probabilities.

Total probability formula: Suppose $\cup_{i=1}^{k} B_i = \Omega, \quad B_i \cup B_j = \emptyset$. (We call this a partition of Ω.)

Then, $P(A) = \sum_{i=1}^{k} P(A \mid B_i) P(B_i)$

Proof:

$$
\begin{aligned}
A &= A \cap \Omega \\
&= A \cap (\cup_{i=1}^{k} B_i) \\
&= \cup_{i=1}^{k} (A \cap B_i)
\end{aligned}
$$

It follows that,

$$
\begin{aligned}
P(A) &= P\left(\cup_{i=1}^{k} (A \cap B_i)\right) \\
&= \sum_{i=1}^{k} P(A \cap B_i) \quad (\text{because } (A \cap B_i) \cap (A \cap B_j) = \emptyset) \\
&= \sum_{i=1}^{k} P(A \mid B_i)\, P(B_i) \cdot \qquad \square
\end{aligned}
$$

Now,

$$\begin{aligned} P(B_i \,|\, A) &= \frac{P(A \cap B_i)}{P(A)} \\ &= \frac{P(A \,|\, B_i)P(B_i)}{P(A)} \\ &= \frac{P(A \,|\, B_i)P(B_i)}{\sum_{j=1}^{k} P(A \,|\, B_j)P(B_j)} \end{aligned}$$

Bayes' Formula (an inversion formula)

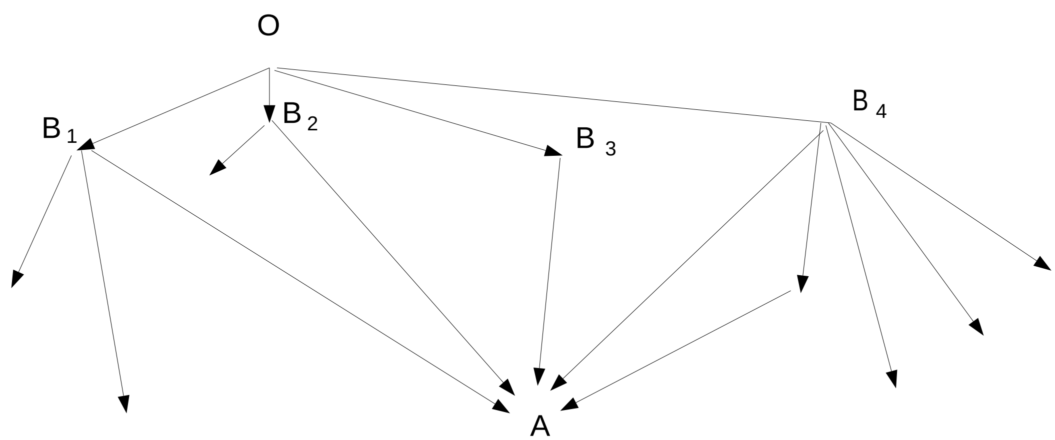
$$\boxed{P(B_i \,|\, A) = \frac{P(A \,|\, B_i)\, P(B_i)}{\sum_{j=1}^{n} P(A \,|\, B_j)\, P(B_j)}}$$

This formula has its origins in two very famous 18^{th} century papers by Reverend Thomas Bayes.

(i) "An essay toward solving a problem in the doctrine of chance," Philosophical Transactions of the Royal Society, 1763, pp 370-418 (reprinted in *Biometrika* 45:293-315, 1958).
(ii) "A letter on asymptotic series ...," Philosophical Transactions of the Royal Society, 1763, pp 269-271.

This is the most important formula in our subject. Various other versions prove to be versatile in telling us how to update or evolve probabilities using data. This is somewhat like Newton's $m\ddot{x} = f$, telling us how to evolve particle motions.

Example (Hiking): Hiker leaves O, choosing one of the roads OB_1, OB_2, OB_3, OB_4 at random. At *each subsequent fork*, he again chooses a road at random. What is the probability of the hiker arriving at point A?


O
B 1
B 2
B 3
B 4
A

$$\begin{aligned}
P(B_k) &= \frac{1}{4}, \quad k = 1, 2, 3, 4 \\
P(A \mid B_1) &= \frac{1}{3} \\
P(A \mid B_2) &= \frac{1}{2} \\
P(A \mid B_3) &= 1 \\
P(A \mid B_4) &= \frac{2}{5} \\
P(A) &= \frac{1}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot \frac{2}{5} \\
&= 67/120 \quad \text{(by total probability formula).}
\end{aligned}$$

We can ask a related question. If the hiker arrives at A, what is the probability that he passed through B_2? This is just $P(B_2 \mid A)$ and Bayes' formula gives us

$$\begin{aligned}
P(B_2 \mid A) &= \frac{P(A \mid B_2)\, P(B_2)}{\sum_{k=1}^{4} P(A \mid B_k)\, P(B_k)} \\
&= \frac{1/2 \cdot 1/4}{67/120} \\
&= \frac{120}{8 \cdot 67} \\
&= \frac{15}{67} \cdot \quad \square
\end{aligned}$$