NMAI059 – Probability and Statistics 1

Mykhaylo Tyomkyn

Lecture 2 - Conditional probabilities. Independence of events.

We want to study interactions between events and their probabilities.

Definition 1 (Conditional probability) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $A, B \in \mathcal{F}$ be events with $\mathbb{P}(B) > 0$. We define the* probability of A conditioned on B *as*

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

Remark 1 *If B is fixed and for all $A \in \mathcal{F}$ we define $\tilde{\mathbb{P}}(A) = \mathbb{P}(A \mid B)$ then $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ defines a probability space.*

Exercise 1 *Prove it.*

Example 1 *To model a roll of two dice[1], we put $\Omega = [6]^2, \mathcal{F} = 2^\Omega$, and $\mathbb{P} = Unif(\Omega)$ – the uniform measure. Let A be the event "The sum the numbers rolled is 8" and let B be the event "the first number is a prime". Then $\mathbb{P}(A) = 5/36$, $\mathbb{P}(B) = 1/2$ and $\mathbb{P}(A \cap B) = 1/12$, so*

$$\mathbb{P}(A \mid B) = \frac{1}{6} \quad \textit{and} \quad \mathbb{P}(B \mid A) = \frac{3}{5}.$$

Definition 2 *Two events $A, B \in \mathcal{F}$ are called* independent *if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Equivalently, assuming $\mathbb{P}(B) \neq 0$, A and B are independent if[2] $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.*

Independence of events often occurs when one would naturally expect it, for instance when considering multiple coin tosses or rolls of dice. Sometimes however, we have independence where 'causal independence' is not apparent.

Example 2 *Consider the probability space of Example 1. Let A be the event "The first roll is a 1" and let B be "the sum of the numbers is 7". Then A and B are independent.*

Theorem 1 (Chain rule) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A_1, \dots, A_n \in \mathcal{F}$ be events with $\mathbb{P}(\bigcap_{i=1}^{n-1} A_i) > 0$. Then*

$$\mathbb{P}(\bigcap_{i=1}^{n} A_i) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid \bigcap_{i=1}^{n-1} A_i).$$

[1] $[n]$ will denote the set $\{1, \dots, n\}$

[2] Or, symmetrically, assuming $\mathbb{P}(A) \neq 0$, if $\mathbb{P}(B \mid A) = \mathbb{P}(B)$

Proof Using the definition of conditional probability the above right hand side can be expressed as

$$\mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \cdot \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} \cdot \frac{\mathbb{P}(A_1 \cap \cdots \cap A_n)}{\mathbb{P}(A_1 \cap \cdots \cap A_{n-1})},$$

which telescopes to $\mathbb{P}(\bigcap_{i=1}^n A_i)$. $\square$

The chain rule is an excellent tool to deal with iterated random experiments, when conditional probabilities present themselves naturally.

Example 3 *We draw* 3 *cards from a standard deck of* 52 *cards, without replacement. [Exercise: define an appropriate probability space]. Let A be the event that no 'hearts' card was drawn. Then $A = A_1 \cap A_2 \cap A_3$ where each A_i is the event "the i-th card drawn was not hearts". So, by the chain rule,*

$$\mathbb{P}(A) = \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}.$$

Theorem 2 (Law of total probability) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{B_1, B_2, \dots\}$ be a finite or countably infinite collection of events forming a partition of Ω, i.e. $\bigcup_i B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for all $i \neq j$. Suppose further that $\mathbb{P}(B_i) > 0$ for all i. Then for any $A \in \mathcal{F}$ we have*

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Proof Since the sets $A \cap B_i$ form a partition of A, using σ-additivity and the definition of conditional probability we obtain

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

$\square$

Corollary 1 *Let B be event with $0 < \mathbb{P}(B) < 1$, and $\overline{B} = \Omega \setminus B$ be its complement. Then for any event A we have*

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid \overline{B})\mathbb{P}(\overline{B}).$$

As an application, let us consider a famous question in probability theory, the so-called *gambler's ruin* problem, also known as the *symmetric random walk* on a path graph.

We have two players, A and B, who have initial capitals of a and b Czech crowns, respectively. The game is played in rounds. In each round a fair coin is tossed. If it comes up 'heads', A receives 1 crown from B, and if 'tails', B receives 1 crown from A. The game continues until one of the players has no money left. That player has lost the game. What are the probabilities of each of the three possible outcomes of the game: player A wins / player B wins / the game lasts forever?

To model the game, we take $\Omega = \{-1, 1\}^{\mathbb{N}}$. Each $\omega \in \Omega$ is an infinite sequence of -1's and 1's, representing the outcome of an infinite sequence of coin tosses (1 is 'heads' and -1 is 'tails'). We take[3] $\mathcal{F} = 2^{\Omega}$, and the probability function $\mathbb{P}$ is defined to satisfy

$$\mathbb{P}(\omega \text{ begins with S}) = 2^{-|S|}$$

[3] Not quite, because Ω is uncountable, but let us work under this simplified assumption.

for any *finite* string of ± 1's ($|S|$ stands for the length of S). It is possible to extend this function to $\mathcal{F}$ essentially uniquely.

Now, for each integer t with $0 \le t \le a+b$ let A_t be the event[4] "player A wins the game starting with a capital of t against player B with a capital of $a+b-t$". Note that $\mathbb{P}(A_0) = 0$ and $\mathbb{P}(A_{a+b} = 1)$. Let C be the event that the first coin toss is 'heads'. Then, by the law of total probability, for all $0 < t < a+b$ we have

$$\begin{aligned}\mathbb{P}(A_t) &= \mathbb{P}(A_t \mid C)\mathbb{P}(C) + \mathbb{P}(A_t \mid \overline{C})\mathbb{P}(\overline{C}) = \mathbb{P}(A_t \mid C) \cdot \frac{1}{2} + \mathbb{P}(A_t \mid \overline{C}) \cdot \frac{1}{2} \\ &= \frac{\mathbb{P}(A_{t+1}) + \mathbb{P}(A_{t-1})}{2}.\end{aligned}$$

The last step is justified by the fact that, if the first toss is 'heads', player A will have a capital of $t+1$ vs. B with a capital of $a+b-t-1$, so A would have to win the game under the conditions of A_{t+1} (as the game has no 'memory'). And similarly if the first toss is 'tails'.

Put $p_t := \mathbb{P}(A_t)$. Then $p_0 = 0$, $p_{a+b} = 1$, and $p_t = (p_{t+1} + p_{t-1})/2$ for all $0 < t < a+b$. Resolving this system of linear equations yields $p_t = t/(a+b)$ for all $0 \le t \le a+b$. In particular,

$$p_a = \frac{a}{a+b} \quad \text{and} \quad p_b = \frac{b}{a+b},$$

resulting[5] in the answer to the original question: the probability that player A wins is $a/(a+b)$, that of player B winning is $b/(a+b)$, and that the game lasts forever is 0.

The next theorem involving conditional probabilities is of great practical importance.

Theorem 3 (Bayes' rule) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $B_1, B_2 \ldots$ be a finite or countably infinite collection of events, forming a partition of Ω, and with $\mathbb{P}(B_i) > 0$ for all i. Let A be an event with $\mathbb{P}(A) > 0$. Then for each j we have:*

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

Proof We have

$$\mathbb{P}(B_j \mid A)\mathbb{P}(A) = \mathbb{P}(B_j \cap A) = \mathbb{P}(A \mid B_j)\mathbb{P}(B_j),$$

proving the first identity. The second one follows by the law of total probability. □

An interpretation of the Bayes rule is as follows. $B_1, B_2, \ldots$ are hidden states of the world. The probabilities $\mathbb{P}(B_i)$ are our prior theory about the world. The event A is a new observation. The conditional probabilities $\mathbb{P}(A \mid B_j)$ are the prior theory about the observation, and the conditional probabilities $\mathbb{P}(B_j \mid A)$ constitute an updated theory, based on the observation.

Example 4 *A genetic test aims to detect a certain genetic defect. It is known that 1% of people have the defect. In those people the test detects it with probability 90% (true positives). In people who do not have the defect, the test 'detects' it with probability 9.6% (false positives). If a person gets a positive test result, what are the odds they actually have the genetic defect?*

[4] Harder exercise: Formalize it as a property of ± 1-strings

[5] By definition, p_b is the probability of A winning with capital b against B with capital a, but swapping the names, it is the same as B winning with capital b vs A with capital a.

With A: "test is positive", and B: "the person has the defect", Bayes' formula gives:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid \overline{B})\mathbb{P}(\overline{B})} = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.096 \cdot 0.99}$$

which is about 8.65%.

Let us now define independence of events in more generality.

Definition 3 *The events $\{A_i : i \in I\}$ (I is a set of indices, possibly infinite) are called* (mutually) independent *if for every **finite** $J \subseteq I$ we have*

$$\mathbb{P}(\bigcap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j).$$

Note that for $|I| = 2$ this corresponds to our previous definition. In general, however, mutual independence is a stronger property than pairwise independence.

Example 5 *Two dice are thrown.*
Event A: "the first roll a 1".
Event B: "the second roll a 1".
Event C: "the sum of the two rolls is 7.
These three events are not independent, but any two of them are.