[STAT-04] Conditional probability

Miguel-Angel Canela
Associate Professor, IESE Business School

Conditional probability

Let A and B be events, with $\mathrm{p}[B] \neq 0$. The probability of A **conditional** to B is defined as

$$\mathrm{p}\big[A|B\big] = \frac{\mathrm{p}\big[AB\big]}{\mathrm{p}[B]}\,.$$

This definition only makes sense when the conditioning event has non-zero probability. We will overcome this limitation through the conditional expectation. Mind that, even if it may look a bit exotic, conditioning to events of probability zero is needed in applications.

For two nonzero probability events A and B, statistical independence is equivalent to $\mathrm{p}[A] = \mathrm{p}\big[A|B\big]$. This provides an intuitive definition of statistical independence: A and B are independent when knowing that B occurred does not change the probability of A.

Example 1. What is the probability that the sum of the outcomes of two dice is less than 8, if we know that it is odd? Let A the event that the sum is less than 8 and B the event that it is odd. The numerator and the denominator in the definition above can be easily calculated by counting cases and dividing by 36. Cancelling out 36, we get

$$\mathrm{p}\big[A|B\big] = \frac{2+4+6}{2+4+6+4+2} = \frac{2}{3}\,. \ \square$$

Conditional probability comes easily when:

- Probabilities are regarded as based on the information available. A frequent question is: What is the probability of A, *knowing that* B has occurred? The answer would be the probability of A conditional to B, which I denote here by $\mathrm{p}[A|B]$.
- Focusing on a subpopulation. Let me consider, again, the event of a newborn being male. If we consider the probability of this event, but conditional to the mother being older than 30, we are restricting the analysis to a subpopulation of the population of births.

Note that, since observational studies are usually restricted to a certain subpopulation (companies big enough, male executives, working wives, etc), considering a probability as conditional is just a matter of convenience. We do it when it is practical to do so.

A conditional probability is itself a probability. That is, if we define $\mathrm{p}_B[A] = \mathrm{p}\big[A|B\big]$, we get a probability p_B on the same set of events. It is easy to understand this probability: (a) any event that does not meet B has probability zero, and (b) for the rest, the probability of A is the probability of the intersection AB, rescaled by dividing by $\mathrm{p}[B]$.

The multiplication rule

From the definition of the conditional probability we get directly the **multiplication rule** (note that the roles of A and B are interchangeable),

$$\mathrm{p}\big[AB\big] = \mathrm{p}[A]\,\mathrm{p}\big[B|A\big].$$

The multiplication rule is extended to n events,

$$\mathrm{p}\big[A_1 \cdots A_n\big] = \mathrm{p}[A_1]\,\mathrm{p}\big[A_2|A_1\big]\,\mathrm{p}\big[A_3|A_1A_2\big] \cdots \mathrm{p}\big[A_n|A_1A_2\cdots A_{n-1}\big],$$

by applying the rule successively to p, p_{A_1}, $\mathrm{p}_{A_1A_2}$, etc.

Example 2. Imagine a box containing 5 red balls and 5 blue balls. If we draw three balls at random, what is the probability that the first two are red and the last one is blue?

For the first ball, the probability of being red is 5/10, for the second ball, given that the first ball is red, the probability of being red is 4/9. Finally, for the third ball, given that the first two have been red, the probability of blue is 5/8. Therefore, the probability sought is the product, 0.139. □

Bayes formula

I present in this paragraph two important formulas. First, for a partition $B_1, \ldots, B_k$ of the sample space, we can write $A = (AB_1) \cup \cdots \cup (AB_k)$ and apply axiom [P3], getting

$$\mathrm{p}[A] = \mathrm{p}[B_1]\,\mathrm{p}[A|B_1] + \cdots + \mathrm{p}[B_k]\,\mathrm{p}[A|B_k].$$

This is the **formula of total probability**. Note that $\mathrm{p}[A]$ appears as a weighted average of the probabilities of A in the k cases defined by the partition. The weights are the probabilities of these cases.

The **Bayes formula**,

$$\mathrm{p}[B|A] = \frac{\mathrm{p}[A|B]\,\mathrm{p}[B]}{\mathrm{p}[A]}\,,$$

is the cornerstone of Bayesian statistics. It follows directly from the definition of the conditional probability. Combining Bayes formula with the formula of total probability, we get

$$\mathrm{p}[B_i|A] = \frac{\mathrm{p}[A|B_i]\,\mathrm{p}[B_i]}{\mathrm{p}[B_1]\,\mathrm{p}[A|B_1] + \cdots + \mathrm{p}[B_k]\,\mathrm{p}[A|B_k]}\,.$$

Example 3. An insurance company has three types of costumers: high, medium and low risk. 20% of the customers have high risk, 30% have medium risk, and 50% low risk. The probability that a customer has at least one accident in the current year is 0.25 for high, 0.16 for medium and 0.10 for low risk. What is the probability that a customer has high risk, given that he/she has had at least one accident during the current year?

I denote by R_1, R_2 and R_3 the three risk groups, and by A having at least one accident. Then,

$$\mathrm{p}[R_1] = 0.2, \quad \mathrm{p}[R_2] = 0.3, \quad \mathrm{p}[R_3] = 0.5, \quad \mathrm{p}\big[A|R_1\big] = 0.25, \quad \mathrm{p}\big[A|R_2\big] = 0.16, \quad \mathrm{p}\big[A|R_3\big] = 0.10.$$

Now, Bayes formula gives

$$\mathrm{p}\big[R_1|A\big] = \frac{0.2 \times 0.25}{0.2 \times 0.25 + 0.3 \times 0.16 + 0.5 \times 0.10} = 0.338. \ \square$$

Conditional independence

As mentioned above, a conditional probability is a probability. When we consider independence with respect to this particular probability, we call it **conditional independence**. So, conditional to C, A and B are (statistically) independent when

$$\mathrm{p}\big[AB|C\big] = \mathrm{p}\big[A|C\big]\,\mathrm{p}\big[B|C\big].$$

The extension to more than two events is done in the obvious way. The following equivalent statements provide different views of conditional independence:

- A and B are independent conditional to C.
- $\mathrm{p}\big[A|BC\big] = \mathrm{p}\big[A|C\big]$.
- AC and BC are independent.

Two events can be conditionally independent without being independent. This may seem, at first sight, an excuse for the professor to introduce an exotic counterexample, but it is not. Many theoretical issues discussed in managerial science papers are concerned with this distinction, which can be seen as a particular case of a more general topic, the **Simpson paradox**. There are many versions of this paradox, always relying on the same fact: conditioning changes, sometimes dramatically, the probabilities.

I restrict the actual discussion to the simplest issue, the distinction between conditional and unconditional independence, illustrating it with a challenging example, which shows that we can take two tosses of one coin as either dependent or independent observations.

Example 4. Being used to take successive coin tosses as statistically independent in elementary probability exercises, we don't bother mentioning it. But it is not so simple, since what we really have is independence conditional to the coin being specified. Let me consider an experiment involving two coins, one with one head and one tail and the other with two heads. I first choose at random the coin, and then I toss twice the selected coin.

I denote by F the choice of the fair coin, and by H_1 and H_2 getting head in the first and second toss, respectively. Applying the formula of total probability to the partition formed by F and F^c, we get

$$\mathrm{p}[H_1] = \mathrm{p}[H_2] = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot 1 = \frac{3}{4}\,, \qquad \mathrm{p}\big[H_1H_2\big] = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot 1\cdot 1 = \frac{5}{8}\,,$$

showing that H_1 and H_2 are not independent. Note that

$$\mathrm{p}\big[H_2|H_1\big] = \frac{\mathrm{p}\big[H_1H_2\big]}{\mathrm{p}[H_1]} = \frac{5}{6}\,.$$

A Bayesian will tell you the story in the following terms. A priori, the probability of getting head in the second row is 3/4. After observing the first toss (a posteriori), the probability is 5/6, higher because I got head in the first row (if I had gotten tail, it would be lower, 1/2). □

Homework

A. In a town of $n+1$ inhabitants, a person tells a rumor to a second person, who in turn repeats it to a third person, etc. At each step, the recipient of the rumor is chosen at random among the other n persons. Find the probability of the rumor being told exactly r times (including the first person telling it to the second person):

(a) Before returning to the originator.

(b) Without being repeated to any person.

B. In a bestselling book, the author reports a test on the ability of German physicians to read the numbers. About 95% of those enrolled in a study were wrong in the following one:

The probability that a woman in a certain population has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90% that she will have a positive mammogram and, if she does not have breast cancer, the probability is 7% that she still will have a positive mammogram. If a woman gets a positive mammography, what is the probability that she actually has breast cancer?

¶ Source: G Gigerenzer (2002), *Reckoning with risk*, Penguin.

C. There are many versions and extensions of the *Monty Hall problem*, all of them descendants of a probability classic, the *Bertrand's box paradox*. Monty Hall was the host on a TV classic called *Let's Make a Deal*. Monty gave the player the choice of three doors: behind one door was a car and, behind the other two doors, goats. The player picked a door, say No. 1, and Monty, who knew what was behind the doors, opened another door, say No. 3, which had a goat. He then said to the player,"Do you want to pick door No. 2?"

Would you switch the initial choice?