Conditional probability (Cont) - Independent events

Math 477

September 26, 2014

1 The law of total probability

In the above section, we have used the formula:

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

This formula can be generalized in the following way: suppose $F_1, F_2, \cdots, F_n$ are mutually exclusive events and $\cup_{k=1}^n F_k = S$ (We say that the events $F_k, k = 1, \cdots, n$ *partition* the sample space) then

$$P(B) = \sum_{k=1}^{n} P(BF_k) = \sum_{k=1}^{n} P(B|F_k)P(F_k). \tag{1}$$

We refer to the above formula as the *law of total probability*.

This law is *very useful* to us because of 2 reasons. We'll list them below.

1.1 Information given in terms of conditional probability

Sometimes information is just given to us in the form of conditional probability. The law of total probability allows us to go from the "conditional form" to the "unconditional form". The following example illustrates.

Example 1.1. *Suppose the probability that an accident-prone person will have an accident within next year is .4, while the probability that a non-accident-prone person will have an accident next year is .2. If 30 % of the population is accident-prone, what is the probability that a (randomly selected) person will have an accident within next year?*

Ans: Note that the probability given in the example are *conditional probabilities*. **If** a person is accident-prone then the probability is $\cdots$. Thus we need to express it properly. Let E be the event that a person is accident-prone and F be the event that a person will have an accident next year. Then

$$\begin{aligned} P(F|E) &= 0.4 \\ P(F|E^c) &= 0.2 \end{aligned}$$

We're asked for $P(F)$. Thus by the law of total probability

$$P(F) = P(F|E)P(E) + P(F|E^c)P(E^c) = (0.4)(0.3) + (0.2)(0.7) = 0.26$$

1.2 Using the law of total probability to divide problems into different stages

The multiplication rule allows us to think about the problem in different natural "stages," as the following example demonstrates.

Example 1.2. *The chess clubs of 2 schools A and B consist of, respectively, 8 and 9 players. 4 players from each club are randomly selected and randomly matched. Suppose that Rebecca belongs to the chess club of school A and her sister Elise belong to the chess club of school B. What is the probability that*

a. Rebecca and Elise will be paired?

b. Rebecca and Elise will be chosen but not played each other?

Answer:

a. P(R and E paired) = P (R and E paired — R and E both selected) P(R and E both selected) + P(R and E paired — either R or E not selected) P (R or E not selected).

But P(R and E paired — either R or E not selected) = 0. Thus P(R and E paired) = P (R and E paired — R and E both selected) P(R and E both selected). Also note that P(R and E paired — Rand E both selected) = 1/4 (looking from either the point of view of R or E). So we only need to compute P(R and E both selected).

You can compute P(R and E both selected) by counting, but we can use conditioning again.

P(R and E both selected) = P(R and E both selected — E selected) P(E selected) + P(R and E both selected — E not selected) P (E not selected).

Again P(R and E both selected — E not selected) = 0. P(R and E both selected — E selected) = P(R selected from her school — E selected) = 1/8. P(E selected) = 1/9. Thus we have our answer.

b. Similar to the above,

P(R and E chosen, not paired) = P(R and E chosen, not paired — R and E chosen) P(R and E chosen) + something with probability 0.

P(R and E chosen, not paired — R and E chosen) = P(R and E not paired — R and E chosen) = 3/4. Thus we also have the answer here as well.

1.3 Computing probability by conditioning

Sometimes it is harder to compute the probability of an event, say $P(A)$, but it may be easier to compute $P(A|F_k), k = 1, \cdots, n$ where $F_k, k = 1, \cdots, n$ partitions the sample space. Thus if we know $P(F_k)$ then using the law of total probability we can compute $P(A)$.

Conditional expectation doesn't have to be complicated. *Actually it can be easier, since when you condition on some event, you generally have more information than you got before.* And so computing probability by conditioning gives us a potentially easier way to handle the problem. We illustrate through the following example.

Example 1.3. *A deck of cards is shuffled, and the cards ared turned up one at a time until the first ace appears. Is the next card - the one following the first ace - more likely to be the ace of spades or two of clubs?*

Ans 1 (Using combinatorics) The key to answering this question is to have the correct "global" point of view (you don't want to consider what the 1st card, 2nd card, 3rd card $\cdots$ could be). You probably think of considering a deck of 48 cards and the number of ways you can insert the remaining 5 cards of interest (4 aces and the 2 of clubs) into the deck. But this is still complicated. There is only ONE card of interest (either the ace of spade or the two of clubs). Thus we look at a deck of 51 cards, and how many ways we can insert the remaining card of interest into the deck so that it's next to the first ace. But there's only 1 way of doing this. So the probability is

$$\frac{51!}{52!} = \frac{1}{52}.$$

Ans 2 (Using conditioning) This answer can also be approached using conditioning, but you want to be careful on what events you condition on. It is because to use the total law of probability, we need to know the probability of the events we condition on. Also we want to make sure that the events we condition on will make it easier to solve the problem (otherwise there's not much point in conditioning on those events).

For example, one might try to condition on the event E_k that the first ace shows up on the kth draw, $k = 1, 2, \cdots, 49$. Let F be the event that the 2 of clubs follow the first ace. Can we compute $P(F|E_k)$ and $P(E_k)$? You may see that $P(E_k)$ is not easy to compute. Also $P(F|E_k)$ is also not easy, (it's not $\frac{1}{51}$) as we need to consider the positions of the remaining aces.

So the correct events to condition on is the ordering of the remaining 51 cards, besides our card of interest. Let us call those events E_j, $j = 1, \cdots, 51!$. Then

$$P(F) = \sum_j P(F|E_j)P(E_j).$$

The reason why conditioning on these events is correct is because given any particular ordering E_j of the 51 cards, $P(F|E_j) = \frac{1}{52}$ (there are 52 "gaps" created by 51 cards, including the 1 before and the one after all the 51 cards. Only 1 gap makes the 2 of clubs next to the 1st ace). Thus

$$P(F) = \sum_j P(F|E_j)P(E_j) = \frac{1}{52} \sum_j P(E_j) = \frac{1}{52}.$$

2 Independent Events

2.1 Independence of 2 events

Definition 2.1. *Two events A and B, are said to be independent if $P(A|B) = P(A)$ and $P(B|A) = P(B)$.*

Remarks: If $P(A|B) = P(A)$ then $P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = P(B)$. Thus we actually need one of the two equalities given above for the definition of 2 independent events.

Interpretation: Intuitively, two events are independent if the knowledge of one event already happened does not influence the probability of the other happening, hence the definition.

Alternatively, one can define A and B to be independent if $P(A \cap B) = P(A)P(B)$. You should check that this is equivalent to the condition $P(A|B) = P(A)$ given in the definition. So in fact one have two possible ways to define what it means for 2 events to be independent. The advantage of using the equality $P(AB) = P(A)P(B)$ is that it is defined even when $P(A)$ or $P(B) = 0$. On the other hand, the interpretation of the equality $P(A \cap B) = P(A)P(B)$ is not as clear as $P(A|B) = P(A)$. Computationally, it is easier to use the definition $P(AB) = P(A)P(B)$ for checking independence.

We have the following easy consequence: If A and B are independent then A and B^c are also independent.

Proof: $P(AB^c) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$.

Note: apply the result again we also have (A^c, B) and (A^c, B^c) are also independent.

2.2 Examples

We very often believe the 2 coin tosses or the 2 die tosses are independent, because "the one toss doesn't influence the other's result." As you see, from a mathematical point of view this explaination is not satisfactory. We really need to check, under a probability distribution (say when the outcomes are equally likely), that the definition of independence given above is satisfied.

Example 2.2. *When a coin is fair, then $P(\text{1st toss is H}) = P(\text{2nd toss is H}) = \frac{1}{2}$ and $P(HH) = \frac{1}{4}$. Thus it is true that $(HH) = P(\text{1st toss is H})P(\text{2nd toss is H})$ and so the events that the 1st toss is H and the 2nd toss is H are independent. But technically we'll need to check for all other similar events to get to the conclusion that the 1st toss's results and 2nd toss's results are independent.*

It is also reasonable to believe that for 3 events E, F, G to be independent, we'll need them to be pairwise independent. That is $(E, F); (F, G); (E, G)$ are independent. This turns out to be not enough, as the following example shows.

Example 2.3. *A fair die is toss twice. Let E be the event that the sum is 7, F the event that the first toss is 4, G the event that the second toss is 3. Then*

$$P(EG) = P(EF) = P(FG) = \frac{1}{36}.$$

It's also clear that $P(F) = P(G) = P(E) = \frac{1}{6}$. Thus they are pairwise independent.

*However, $P(E|FG) = 1 \neq P(E)$. Thus E and FG are **not** independent, which is a reasonable thing to expect if we say E, F, G are independent.*

2.3 Independence of more than 2 events

Definition 2.4. *The events $E_1, E_2, \cdots, E_n$ are independent if for any subset $\{i_1, i_2, \cdots, i_k\}, 1 \leq k \leq n$ of $\{1, 2, \cdots, n\}$ we have*

$$P(E_{i_1}E_{i_2}\cdots E_{i_k}) = P(E_{i_1})P(E_{i_2})\cdots P(E_{i_k}).$$

In particular, when $n = 3$ we say E, F, G are independent if

$$\begin{aligned} P(EFG) &= P(E)P(F)P(G) \\ P(EF) &= P(E)P(F) \\ P(FG) &= P(F)P(G) \\ P(EG) &= P(E)P(G). \end{aligned}$$

3 Specifying independent events

There are cases when a problem specifies a priori that events are independent. Consider the following example.

Example 3.1. *We toss a coin, whose probability of head is 1/3 and tail is 2/3 three times. Assuming the results are independent among the tosses. What is the probability of getting a specific sequence, i.e. HHH or HTH etc? Ans: By the independence assumption, $P(HHH) = \big(1/3\big)^3$ and $P(HTH) = \big(1/3\big)^2 2/3$.*

Note that the above example can be equivalently formulated as

Example 3.2.

$$\begin{aligned} P(HHH) &= (1/3)^3, P(TTT) = (2/3)^3, \\ P(HTH) &= P(THH) = P(HHT) = (1/3)^2 2/3, \\ P(TTH) &= P(THT) = P(HTT) = (2/3)^2 1/3 \\ \cdots \end{aligned}$$

Then the one of the thing to show is the results among the tosses are independent among the tosses. This cannot be assumed apriori in the formulation of the second example.

There are very easy examples where we give a distribution and the results among trials are not independent. Consider the following example.

Example 3.3. *Suppose we toss a coin twice and we have*

$$\begin{aligned} P(HH) = P(TT) &= 1/6; \\ P(HT) = P(TH) &= 2/3. \end{aligned}$$

Then you see that $P(1stH) = 1/2$ *and* $P(2ndH) = 1/2$ *but* $P(HH) = 1/6$ *so the 1st and 2nd toss are not independent.*

So we should view specifying independent events as another way to specify distribution among trials, especially when we don't have reasons to believe results of trials influence one another. This is a very convenient way, when the number of trials are not given beforehand (i.e. we keep tossing coins until seeing the first H). This is the approach of the *problem formulator, not problem solver.* On the other hand, if we're solving a problem in which the independence assumption is not given, just because we believe results of trials are independent does not mean we can use the independence property. We need to show it before we want to use it.