EVENTS -- useful examples in more detail

I. There is a probability example in the text book (see page 109) concerning two flips of a coin chosen at random from two coins, one fair and one biased. Here we will elaborate on that presentation, taking a closer look at the underlying elementary events.

Here is what is given: the ``fair'' coin has equal chances of coming up heads and tails, but the biased coin always comes up heads. Let us use F to stand for the fair coin being the one chosen, and U for the biased (unfair) one being chosen. Then Pr[F] = Pr[U] = 0.5, since one of the two coins is picked at random. As for the probabilities of one of them coning up heads, we need to look at conditional probabilities. For this, it is helpful to write down the elementary events associated with this problem.

An ``experiment'' here involves three things: choosing one of the two coins at random, flipping it once, and flipping it again. The outcomes for the entire experiment then amount to a triple as follows:

<F or U , H or T, H or T>

If we write out all the possible outcomes that can arise from repeating the experiment over and over we get the following, which then are the elements of the sample space S for this problem:

<F,H,H>
<F,H,T>
<F,T,H>
<F,T,T>
<U,H,H>

Note that triples such as <U,H,T> are *not* possible since the biased coin never comes up tails. Thus there are five elementary events; but they are not equally likely. For we are given that F and U are equally likely, so the *non-elementary* event F (the fair coin is chosen) is equally as likely as the event U (the unfair coin is chosen): Pr[F] = Pr[U] = 0.5 as we noted above.

But F, being a non-elementary event, is a subset of S, and in fact it is simply

(1) F = {<F,H,H> , <F,H,T> , <F,T,H> , <F,T,T>};

that is, F consists of all elementary events in which the fair coin is chosen. (And U of course is U = {<U,H,H>}.)

What are the probabilities of the four elementary events that make up F? Well, being fair means that T and H are equally likely for that coin, so we can draw the following table:

```
          |   <F,H,H>     these four outcomes are equally
          |               likely since this coin is fair;
          |   <F,H,T>     this means that each has the
Pr = 0.5 --|               same probability as the other
          |   <F,T,H>     three, and since their sum must
          |               be 0.5 then each must have prob
          |   <F,T,T>     of one fourth of 0.5, i.e., 1/8.


Pr = 0.5      <U,H,H>
```

The above table reveals in detail just what is what. Each elementary event is now given a precise probability, and this will allow us to calculate all the conditional and other probabilities we may be interested in. We have:

```
Pr[<F,H,H>] = 1/8
Pr[<F,H,T>] = 1/8
Pr[<F,T,H>] = 1/8
Pr[<F,T,T>] = 1/8
Pr[<U,H,H>] = 1/2 = Pr[U]
```

For instance, we can calculate Pr[U | HH] -- the conditional probability of U given HH -- as follows:

```
                      Pr[U int HH]          1/2              1/2
(2)    Pr[U | HH] =  --------------  =  -------------  =   -------  = 4/5
                         Pr[HH]          1/8  + 1/2          5/8
```

because (from our table) we see that the event (U intersect HH) is simply the event <U,H,H>: U only appears one time in the table, and both flips are heads for it, and its probability is 1/2; this gives the numerator above. The denominator requires us to find all elem events having HH, and there are exactly two, one with probability 1/8 and one with 1/2, so their sum is the probability of the event HH occurring.

We also can see the same answer more quickly from the table: HH occurs once with prob 1/8 and once with prob 1/2 = 4/8; only the latter value (4/8) has U in it, so of the total of 5 eighths shown, 4 of them have U, which is 4 out of 5, or 4/5. The calculations above amount to exactly this, but in boring detail.

What about Bayes' Theorem? We can use it too, to get the same result in a more involved way:

$$\Pr[U \mid HH] = \frac{\Pr[U] \Pr[HH \mid U]}{\Pr[HH]} = \frac{0.5 * \Pr[HH \mid U]}{5/8} = \frac{0.5 * 1.0}{5/8}$$

$$= 4/5 \quad \text{as before.}$$

How do we know Pr[HH | U] = 1.0 above? Intuitively we know this since we are told the unfair coin always comes up heads. But on a more precise level, we can calculate this from the table, just as we did a little earlier: there is exactly one elem event in which U holds, and HH also holds in that event. Thus (giving all details)

$$\Pr[HH \mid U] = \frac{\Pr[\text{ HH int U}]}{\Pr[U]} = \frac{0.5}{0.5} = 1.0$$

or more simply, HH is true in all events in which U is true, so given that U holds, HH is guaranteed. But the table gives the underlying reasoning that shows this intuition to be correct.

So, Bayes' Theorem is not necessary to do this problem, and in fact it a bit more work than a straightforward solution.

Notice that key to this problem is the recognition that the elem events are actually triples, made up of more complex events. Thus the elem outcome <F,T,H> has ''in'' it the complex event F, the complex event H, and so on. But the elem events are not sets, they are ordered triples. The complex events, such as F, are sets as we saw above.

++
++

II. An interesting additional problem is presented in exercise 6.2-10 (on page 110): behind one and only one of three curtains is hidden a prize. After you have pointed at one curtain, someone raises one of the other two, showing there is no prize behind it. You then are given the choice of staying with the curtain you originally pointed at, nor switching to the third one (that was not lifted). What is the probability the prize is behind your original curtain, and what is the probability it is behind the third curtain?

This is a famous puzzle that you may have come across elsewhere. Many people, including sophisticated statisticians, often get this one wrong when they use intuition. But a careful analysis with the tabular method above makes it clear and also shows how to adjust our intuitions so that the correct answer also eventually *seems* correct!

Here we go: The elementary events are made up of a pointing (by you) at one of the three curtains; and by the raising (by someone else) of another curtain that does *not* have the prize behind it. Let's write A, B, and C for the three curtains. Then exactly one of A, B, C, hides the prize.

Suppose you pick curtain A; if you instead pick B, or C, the analysis is the same. Below is the table of elem events (assuming you pick A, so that the only remaining outcome has to do with which curtain is lifted and which has the prize -- one could write out all the cases of your pointing at A, B, and C of course, but this makes the table three times as long and all three parts will involve the very same analysis, so we'll just do one: case A). We need to specify which curtain has the prize and which is lifted; let's use the curtain name (A,B,C) to indicate that is has the prize, and lowercase (a,b,c) to indicate it is lifted. Then there are four elem events in which we have picked curtain A:

```
prize        lifted      event   probability

  A            b    ]
                    ]--    A        1/3
  A            c    ]


  B            c           B        1/3  ]
                                         ]--2/3
  C            b           C        1/3  ]
```

Note that if B has the prize, then C will be lifted (c is true) for sure, since it is given in the problem that the lifted curtain does *not* have the prize (we can imagine, for instance, that the person who does the lifting knows where the prize is) and is *not* the one we picked. (If we wrote out all the cases, not just that of our picking A, we'd get not four but twelve elem events).

So, the sample space is: S = {<A,b> , <A,c> , <B,c> , <C,b>}.
Let's abbreviate this as S = {Ab,Ac,Bc,Cb}.

What are their probabilities? Since we are only told that the prize is behind one curtain, then we have no reason a priori to prefer one curtain over another: they are all equally likely to hide the prize. So in particular, curtain A has an initial probability of 1/3 of hiding the prize; so does B, and so does C:

Pr[A] = Pr[B] = Pr[C] = 1/3

and also event B = Bc, event C = Cb, event A = {Ab,Ac} (recall that we are using uppercase (A,B,C) to indicate the prize behind the curtain). If A has the prize then either of B or C can be lifted, but if either B or C has the prize then there is only one curtain that can be lifted. We do not know the precise probabilities for elem events Ab and Ac; maybe the person who does the lifting prefers b to c, for instance; but this will not matter: their *sum* is 1/3. (The results below will not depend on knowing these probabilities, but just for concreteness' sake, let us suppose that Pr[b] = Pr[c] = 0.5; that is, the curtain-lifter has no bias about which curtain is lifted when neither has the prize. Then Pr[Ab] = Pr[Ac] = 1/6.)

Now, what is the event whose probability we wish to compute? We want to know the probability that A has the prize, given that a different curtain has been lifted. That is, we want Pr[A | b or c]. This we can calculate:

```
                    Pr[A int (b or c)]       Pr[A]       1/3
Pr[A | b or c]  =  --------------------  =  -------  =  -----  =  1/3
                       Pr[b or c]            1.0         1.0
```

(Note that (b or c) holds in every elem event.)

So, learning that the prize is *not* behind the lifted curtain does not alter the original probability Pr[A] = 1/3.

Now what about switching our choice to the third curtain -- the one (either B or C) that was not lifted? One might guess that it too still has prob 1/3 of having the prize, and that there is no point in switching. But consider two thoughts:

1. the *sum* of the probabilities for our curtain (A) and the unlifted curtain to have the prize must be 1.0 (the prize has to be somewhere, and it is not behind the lifted curtain), and if it is 1/3 for A then

it must be 2/3 for the other one, so we should switch to it.

2. once we learn the prize is not behind the lifted curtain, then it must be behind A or the unlifted curtain (B or C) and so the prob for either of these ought to be 0.5 (so there is no point in switching).

Thus intuitions can be conflict with each other, and we must compute rather than simply guess. Now, (2.) above presumably is wrong, because we already did compute Pr[A | b or c] = 1/3. But let us verify the ideas in (1.) by computing Pr[U], where U stands for the prize being behind the unlifted curtain:

$$\begin{aligned}
\Pr[U] &= \text{sum of probs of all elem events in which U holds} \\
&= \text{sum or probs of elem events in which A does not hold} \\
&= \Pr[B] + \Pr[C] = 1/3 + 1/3 = 2/3
\end{aligned}$$

(Note that U holds iff A does not hold.)

So, it is in fact wise to switch our choice of curtain from A to the unlifted one (B or C), since that doubles our chanced of getting the prize, from 1/3 to 2/3.

Just to get more experience, let us compute further. Suppose curtain C is lifted (c is true). Then

$$\Pr[B \mid c] = \frac{\Pr[B \text{ int } c]}{\Pr[c]} = \frac{\Pr[B]}{0.5} = \frac{1/3}{0.5} = 2/3$$

Or the same result from Bayes:

$$\Pr[B \mid c] = \frac{\Pr[B]\,\Pr[c \mid B]}{\Pr[c]} = \frac{1/3 * 1.0}{0.5} = 2/3$$

because given B, c is guaranteed (Pr[B | c] = 1.0).

One can think about the problem in another way. Suppose instead of three curtains, there are 100, and you choose one. Then of the other 99 curtains, 98 are lifted, all with no prize. Would you switch to the one unlifted curtain of the 99, or stay with your first choice? It is very unlikely that your first choice has the prize (prob = 1/100), so it is very likely (prob = 1 - 1/100 = 99/100) that that other curtain has the prize.

Why? It is because the curtain-lifter knows where the prize is, and is deliberately lifting some of those that do not have the prize. Thus every time a new curtain is lifted, you are being given useful information -- unless of course your curtain had the prize all along, but that is very unlikely. So you should switch, and improve your chances from 0.01 to 0.99!

We can now restate the original problem of three curtains in somewhat more intuitive terms: You pick one curtain, say A, to start with, and you know it has only a 1/3 chance of hiding the prize, and therefore that there is a 2/3 chance it does *not* have the prize (i.e., that either B or C has the prize). But switching away from A is not useful since you would not know whether to switch to B or to C, and each also has the same chance (1/3) as does A. But when (say) C is lifted, you now know it does not have the prize, so now B looms up as a very good choice: it is as is you were told that the 2/3 chance of (B or C) is now all given to B.

Returning one more time to our table, but with a few more details added, especially the assumption that the choice of curtain to lift is made randomly when either of B or C can be lifted (i.e., when neither has the prize):

prize	lifted	elem event	probability	complex event	probability
A	b	Ab	1/6	A	1/3
A	c	Ac	1/6		
B	c	Bc=B=c	1/3	(B or C)=(b or c)	2/3
C	b	Cb=C=d	1/3		

We now can simply read off anything we want to know. For instance, the chance that B has the prize given that C does not (i.e., given that C is lifted, or c holds) simply directs our attention to the rows in which c holds (there are only two, the second and third rows) and in one of them (the second) B does not hold (rather A does) and in the other (the third0 B holds. So, considering only those two rows (shown again below)...

A	c	Ac	1/6

```
        B              c          Bc=B=c     1/3
```

...B has twice the chance of being true as it does of being false (1/3 vs 1/6), so it's chance of holding, GIVEN that c holds (i.e., given that one of these rows is true), is 2 out of 3, or 2/3.

This visual analysis of the table is exactly what the formulas for conditional and other probabilities compute, and in the same way: by restricting the computation to relevant rows, and multiplying or dividing by a suitable factor (in this case, multiplying by 2) so that the total probability remains 1.0. That is, in this example, the shown elem events have total prob of 1/6 + 1/3 = 1/2, so we divide by 1/2 (or multiply by 2). That is what we did intuitively when we reasoned that 1/3 is twice a large as 1/6, so the case of 1/3 corresponds to being twice as probable as 1/6, i.e., the 1/3 case gets 2/3 of the probability! This odd jumping back and forth comes about because we are looking at only some of the rows, and so the prob values do not sum to 1.0; this also is what the prob-formulas do automatically.

It may seem that visualization is faster and easier and more easy to understand, and in fact it is a very good approach to use when the examples are fairly small; but when the table gets large, it is too easy to overlook something, and better to use the formulas. However, the meaning of the formulas is given by what we have just done visually.