Classroom Notes

Math 394B Summer 2005

Week 3

1 Example 3m (p. 81)

The example in question seems to be explained in a less than clear way in the book, so here's another discussion.

1.1 Basic Probabilities

We denote by G (generically) the event that a given individual is guilty , and by M that the DNA test finds a match (appending subscripts to indicate which specific individual we are talking about). The assumptions are that, for any specific individual, the following probabilities hold.

1. For each of the N (=990,000) non-convict residents, $P[G] = \beta$
2. For each of the n (=10,000) released convicts, $P[G] = \alpha$
3. For any person, $P[M|G] = 1, P[M|I] = 10^{-5}$

The probabilities in 1. and 2. are not conditional probabilities: the idea is that we have 100,000 possible events (individual i is guilty), and each of these disjoint events has probability αor βdepending on whether the individual in question belongs to one or the other group. Hence,

$$10000\alpha + 990000\beta = 1$$

$$\beta = \frac{1 - 10^4\alpha}{9.9 \cdot 10^5}$$

Since in this problem only the ex-convict population is tested, the only role of β (and of the general population) is to put a limit on α (we need $\alpha \geq 0$, and $\beta \geq 0$).

1.2 Some Consequences

There are a few events that are observed. First of all, individual #1 (AJ in the book) matches, and 9,999 other ex-convicts do not. Thus we observe $M_1 \cap$

$\bigcap_{j=2}^{10000} M_j^c$ (the index identifies the individual, AJ being #1, and the other ex-convicts going from 2 to 10,000).

First, we may observe that we now know that none of the other ex-convicts is guilty (only innocents fail the matching tests, though they may erroneously match). Note that if $\alpha = 1$, that would settle the case. Assuming $\alpha < 1$ (otherwise, there is no problem), and also $\alpha > 0$ (otherwise, AJ would automatically be presumed innocent - and why test the ex-convicts anyway?), we have that

$$P\left[G_1 \middle| M_1 \cap \bigcap_{j=2}^{10000} M_j^c\right] = \frac{P\left[M_1 \cap \bigcap_{j=2}^{10000} M_j^c \,|G_1\right] P\left[G_1\right]}{P\left[M_1 \cap \bigcap_{j=2}^{10000} M_j^c\right]} =$$

$$= \frac{P\left[M_1 \cap \bigcap_{j=2}^{10000} M_j^c \,|G_1\right] P\left[G_1\right]}{P\left[M_1 \cap \bigcap_{j=2}^{10000} M_j^c \,|G_1\right] P\left[G_1\right] + P\left[M_1 \cap \bigcap_{j=2}^{10000} M_j^c \,|G_1^c\right] P\left[G_1^c\right]}$$

with $P\left[G_1\right] = \alpha$.

1.3 Evaluating Conditional Probabilities

Now, we have to compute the probabilities of our observation, conditional on the two possible cases for our #1 individual.

First, $P\left[M_1 \cap \bigcap_{j=2}^{10000} M_j^c \,|G_1\right]$. Since we are now in the case #1 is guilty, the observed outcome is the combination of the guilty party matching the test, and 9999 innocent parties not matching it. Since each test is supposedly independent, we have a conditional probability of $\left(1 - 10^{-5}\right)^{9999}$

Now, $P\left[M_1 \cap \bigcap_{j=2}^{10000} M_j^c \,|G_1^c\right]$. Here, we still have the factor $\left(1 - 10^{-5}\right)^{9999}$, since the 9999 ex-convicts are necessarily innocent (they didn't match the test), and did not match the test, with the additional factor 10^{-5}, for #1 matching the test, while innocent, times the probability that indeed everyone of the 10000 ex-convicts is innocent, i.e. $1 - 10^4\alpha$.

1.4 Wrapping Up

Combining all the above, the conditional probability we are looking for is

$$\frac{\alpha\left(1 - 10^{-5}\right)^{9999}}{\alpha\left(1 - 10^{-5}\right)^{9999} + \left(1 - 10^4\alpha\right) 10^{-5}\left(1 - 10^{-5}\right)^{9999}} =$$

$$\frac{\alpha}{\alpha + 10^{-5}\left(1 - 10^4\alpha\right)}$$

which is the solution formula.

2 A Recap Of Basic Formulas

As we close the preliminary part of our course, the one devoted to events, rather than random variables, it pays to summarize the basic tools that we gained - and that we will be using over and over in the following weeks.

2.1 Basic Basics

We always work with a set - the "sample space", denoted by Ω, and a selected family of subsets of Ω, the family of "events", denoted by $\mathcal{F}$, with certain properties:

1. If $A_1, A_2, \ldots$ are events, then so is $\bigcup_i A_i$
2. If $A_1, A_2, \ldots$ are events, then so is $\bigcap_i A_i$
3. If A is an event, so is $A^c = \Omega \setminus A$
4. $\Omega$and $\emptyset$are events

In the first two statements, we are considering either a finite or countable collection of events.

The final basic element is a function P whose domain is given by $\mathcal{F}$ (note that its domain is *not* Ω: it is a function defined on sets, not on points), taking real values, between 0 and 1, with the properties

1. For any event A $0 \leq P[A] \leq 1$
2. $P[\Omega] = 1$
3. For any finite or countable collection of events A_i,$P[\bigcup_i A_i] \leq \sum_i P[A_i]$
4. If in the previous point, the events A_i are all disjoint, then the inequality becomes an equality.
5. For any event A, $P[A^c] = 1 - P[A]$ (follows from 2 and 4, since $A \cup A^c = \Omega$, and $A \cap A^c = \emptyset$)
6. $P[\emptyset] = 0$ (follows from 2 and 5).
7. $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
8. $P[A \cap B] = P[A] + P[B] - P[A \cup B]$
9. $P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B] - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C]$
10. (and so on, for 4, 5, ... events: alternatively add and subtract the longer and longer intersections)

As an axiomatic system, we only need the properties 1 and 2 for $\mathcal{F}$, and properties 1, 2, and 3 for - everything else is a theorem.

2.2 Conditional Probabilities

The axioms in the preceding section define what an analyst would call "a (finite, positive) measure space", and could be the start of an abstract analysis course. However, mathematical systems are define by the questions they address, as much as by their axioms. The questions we address in probability have led to the introduction of definitions that gain their usefulness in the specific areas of probability and statistics. First and foremost, is the idea of conditional probability, and, as immediate corollary, of independence.

2.2.1 Definitions

We define

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

It is assumed $P[B] > 0$, otherwise the notion is left undefined[1].

If $P[A|B] = P[A]$, we say that the two events are independent.

Rewriting the definitions, we get useful variations:

1. $P[A \cap B] = P[A|B]\,P[B]$
2. $P[A|B]\,P[B] = P[B|A]\,P[A]$
3. $P[A|B] = \frac{P[B|A]P[A]}{P[B]}$ (Bayes' Rule)
4. A and B are independent if $P[A \cap B] = P[A]\,P[B]$

Note that independence is a symmetric relationship, even if the original definition is apparently not. Also, 4 is normally used as the definition of independence, so that we can include events of probability zero in the notion (even if trivially), so that an event of probability zero is independent of any other event.

Almost from the definition, we instantly get that, if A and B are independent, so are A and B^c, and hence so are A^c and B, and A^c and B^c. Hence, Ω is also (trivially) independent of any other event.

2.2.2 Consequences

The rewriting 1 in the preceding subsection indicates how conditional probabilities help in computing probabilities of intersections. One of the main consequences of this observation is the following

[1] We will see that, apparently at least, we will eventually condition on events of "zero probability". However, this will happen in a very specifically controlled way, and, if analyzed deeper, does not really mean that one is "assuming that an event of zero probability has occurred" at all. Ignoring the distinction can lead to huge messes - and be reassured that there are plenty of respected professors that have fallen victim to this mistake.

Theorem (*Total Probabilities*): Suppose $\{A_i\}$ is a (finite or countable) collection of events, forming a *partition* of Ω (i.e., they are all disjoint, and $\bigcup_i A_i = \Omega$). Then, for any set A,

$$P[A] = \sum_i P[A\,|A_i]\,P[A_i]$$

Proof Follows immediately from point 1 in the preceding section, once we realize that

$$A = \bigcup_i A \cap A_i$$

One situation where this formula gets used over and over again is when we apply Bayes' Rule. In fact, here is a standard scenario:

1. We are trying to ascribe an effect to one of many possible causes (e.g., a symptom to an illness). Let E be the event "the effect has been observed", and E_i be the event "the effect is due to cause i".

2. We know the following: $P[E\,|E_i]$ (how likely it is that cause $\#i$ will indeed be the cause of the effect), and $P[E_i]$ (how likely cause $\#i$ actually may occur).

3. After observing the effect, we "update" our estimate of the probability of each effect through Bayes' Rule.

Bayes' Rule requires an additional input: $P[E]$. And here's where the Total Probabilities Theorem kicks in:

$$P[E] = \sum_i P[E\,|E_i]\,P[E_i]$$

Hence, the "updating" formula will read

$$P[E_k\,|E] = \frac{P[E\,|E_k]\,P[E_k]}{\sum_i P[E\,|E_i]\,P[E_i]}$$