Blood Groups

Thomas Forster

August 15, 2005

The diagram I supplied in the coursework is exactly the diagram supplied me (indirectly) by the blood bank people in one of the Auckland hospitals. What information can we extract from this picture? Why are we trying to extract information at all?? Well, I think this chore is quite a realistic one. A working computer scientist or programmer is liable to be given data in all sorts of formats and asked to process it in ways that the presented format might not invite. Ascertaining which data structure best represents/captures the material your employer or client wants you to think about is something you are going to have to get used to.

Enough of the sermon: back to blood groups. Remarkably the relation of potential-donor is a partial order, as you can tell if you look at the table long enough. I was cued into this because the crosses in the diagram fill it in to make it look like a Sierpinski sponge. You may not wish to know what a Sierpinski sponge is, and i am not going to twist your arm to find out if you have no inclination. What this did for me was make me think that the binary relation probably had some nice cute combinatorial property or other so i set out to ascertain which. It turned out that it was a partial order. (It might be that if you draw a diagram like this for a partial order it always looks like a Sierpinski sponge; i don't know: if i ever have a spare moment—when i retire perhaps—i might check.)

The obvious thing to do with a partial order is to write out a Hasse diagram. This is a much cuddlier way to visualise a partial order than a matrix is. What do we find when we draw the Hasse diagram? We get something that looks like a cube. You know that the cube is a three-dimensional object, but you probably wouldn't on that basis jump to the conclusion that this picture is something to the power three. That would be extremely fanciful. But in this case it turns out to be the right thing to do. That is why i gave you the hint that the power set of $\{0, 1, 2\}$ (or for that matter the power set of any set-with-three-members) is isomorphic to the product of three copies of the poset with two elements. I think the example i gave was the poset $\{0, 1\}$ with $0 < 1$. If you draw the Hasse diagrams of these things you will see they are the same. Why does this help?

It suggests that the cube shape you get is the result of three variables that can each take two values. In this setting this is obviously three genes (*loci*) each of which can take two values (*alleles*). Let these three loci be X/x, Y/y and Z/z. (This notation means that the two alleles at each of these three loci are respectively X and x, Y and y, and Z and z. For some reason when there are two alleles at a locus it is standard to denote one with capital letter and the other with the corresponding lower-case letter.) When we write XyZ for a genotype we mean that a chap with XyZ has allele X at the first locus, y at the second and Z at the third. Let us write the "higher" allele at each locus with a capital letter. That is to say, we have an ordering at each locus: $x < X$, $y < Y$ and $z < Z$.

We now find that the product ordering on the triples will turn out to look just like the cube you get from the ordering on blood groups! For example we have $xyZ \leq XYZ$ beco's XYZ has upper case letters at all places. $xYy \leq xYZ$ beco's the formula on the right has upper case letters wherever the thing on the left does. And so on.

This correspondence between the two decorated cubes will tell you which blood group is which triple. That is to say, each blood group is going to correspond to a triple, and we will assume that the ordering on the triples arises as the product of the three orderings, one at each locus. In effect we are assuming that you can give blood to someone as long as you never have a lower-case letter where they have an upper-case letter. Why might this be? I don't know, since this is all guesswork and all i have to go on is the diagram. It *might* be that—at each locus—if you have an lower-case allele then you produce a protein that is toxic to people who don't produce it. That would explain why XYZ can donate to xyz but not *vice versa*. But it's pure guesswork. I will return to this later.

Two parenthetical points here:

1. Notice that this ordering is nothing to do with dominance of z of Z or *vice versa*.

2. It's true that most (tho' not all) metazoans have *two* copies of each gene (they are *diploid* not *haploid*) so that if you have XyZ on one chromosome and yz on the other, whether you look like XyZ or Xyz depends on which of z and Z is dominant over the other. For the moment it's simplest to pretend that we just have one copy of each gene and worry about the diploidy later.

How are we to match up the vertices on the one cube with the vertices on the other? Well, the two cubes are two Hasse diagrams and we obviously have to identify the top point of one with the top pont of the other: $O-$ is the universal donor and must be XYZ: XYZ is the top triple in the product order on triples and $O-$ is the top blood group. The two bottom elements similarly: $AB+$ is the universal recipient and must be xyz. Groups that can donate to four groups must be xYZ, XyZ and XYz, so $O+$, $B-$ and $A-$ must be XYz XyZ and xYZ, but we can't say which is which. (that is to say, we know we have three loci but at this stage we can't tell them apart) Let us set $O+$ to be XYz and $B-$ to be XyZ. So $A-$ must be xYZ. This means $O+$ can give blood to xYz and Xyz. In fact it can give blood to $A+$ and $B+$, so they must be xYz and Xyz (but not necessarily in that order). $B-$ is XyZ and so can give blood to xyZ and Xyz, and we are told it can give blood to $B+$ and $AB-$. But we are told that the only group (apart from $AB+$) that can receive blood from both $O+$ and $B-$ is $B+$, so $B+$ must be Xyz. This tells us that $AB-$ is xyZ and that $A+$ must be xYz. (I won't ask you to calculate how many ways there are of pairing up the vertices of one cube with the vertices of the other, but there is no doubt that it would make a useful exercise. Had I been able to devise this coursework like a treasure hunt I would have put this in as one of the stages!)

Thus

$O-$	is	XYZ
$O+$	is	XYz
$B-$	is	XyZ
$A-$	is	xYZ
$B+$	is	Xyz
$A+$	is	xYz
$AB-$	is	xyZ
$AB+$	is	xyz

Notice that this fits our rule: You can give blood to another group as long as it does not have an upper case letter in a locus where you have a lower case letter. When i say that $AB+$ is xyz this is of course shorthand for "You are $AB+$ if the three alleles you express are x, y and z" This means that you might be $AB+$ if you are xyz on one chromosome and xyZ on the other, if z is dominant over Z. We don't know yet.

Notice that we have got this far without trying to extract any information from the letters 'A', 'B' and 'C' or the + sign. If we are right in our hunch that there are three loci each with two alleles then we would expect these symbols 'A', 'B', 'O', '+' and '-' to spring into life and start making sense. In fact we now notice that our Z/z locus corresponds to the plus sign: if you are a group ending in '+' then you have z at the third locus, and if you are a group ending in a '-' then you have a Z at the third locus. God knows what the +/- thing means (something to do with monkeys i think) but this does at least suggest that we are on the right track. Come to think of it it also looks to me as if whenever a blood group has an 'A' in its name than it has x rather than X and whenever a blood group has a 'B' in its name than it has y rather than Y. This is looking very hopeful.

(If we'd matched up the vertices differently—as we could have done—we would then find that it' not the z/Z locus that corresponds to +/- but tha x/X or the y/Y.

Al of this was done without taking diploidy into account. What is diploidy? Diploid organisms (like us) have *two* copies of each gene. So if you have one copy that is Z and one copy that is z what happens? Well you might behave as if you had z or you might behave as if you had Z. If the first then z is dominant, and if the second, then Z is dominant. In our discussion above, if all the upper-case letters were dominant (and i think the convention is that the dominant allele is always written in upper case) then you would have an $O-$ blood type if you had one copy that was XyZ and the other that was xYz. I think it's quite clear that we can say nothing about which alleles are dominant—after all we have no information about how the groups are inherited. But we can make a few safe predictions.

Is X dominant or is x dominant? Is Z dominant or is z dominant? Is Y dominant or is y dominant? We have eight possible combinations, and we have no way of inferring from the table which combination is the true one. This is because the table provides no information about what flavour of parent can have what flavour of child. If the upper-case alleles were dominant then all the combinations XYZ/xyz, XyZ/xYz, XYz/XyZ and so on—where every uppercase letter appears in one triple or both— would be $O-$ and the only way you could be $AB+$ would be to be xyz/xyz. We simply can't tell from the diagram: it doesn't contain that information. (This being a discrete maths course i think it is fair to ask you—on the assumption that the upper case letters are dominant—how many combinations (like XYZ/xyz, XyZ/xYz, XYz/XyZ and so on) give you which blood group. **Bonus/consolation points will be awarded to teams supplying correct answers**.

What can we do with what we've got? These eight possibilities can be settled by three yes/no questions, and I think these three will settle it

1. Can two $O-$ parents produce an $O+$ baby?
2. Can two $O-$ parents produce a $B-$ baby?
3. Can two $O-$ parents produce an $A-$ baby?

I think if I'm told the answers to those three questions I will be able to give answers to all questions of the form "Can parents of this blood type and that blood type produce a child of that other blood type?". That's beco's—if i've done my sums right—

1. if Z is dominant over z then two parents that were both XYZ/XYz (and therefore $O-$) could produce an XYz/XYz baby, which would be $O+$. Similarly
2. if Y is dominant over y then two parents who were both XyZ/XYZ (and were therefore both $O-$) could produce an XyZ/XyZ baby which would be $B-$.
3. is analogous.

One last thing. I guessed earlier (page 2) that it just might be that people with lower-case alleles produce a protein that is toxic to people with the corresponding upper-case allele, which would be why xyz is the universal recipient but cannot donate to anyone with an upper case allele. If that is so then all the lower case alleles would be dominant: if you are xYX/XYZ then your body produces chemicals as if you were xYZ. This should mean that being XYZ (which is $O-$) is recessive; the only way you can be $O-$ is to be XYZ/XYZ so two $O-$ parents can only produce $O-$ children. Apparently this is true. This doesn't mean that my guess is right, but it looks good.

Please get back to me if there are aspects of this handout that you don't get. It should be intelligible to someone (such as your humble correspondent) who knows nothing whatever about blood groups and knows only secondary-school basics about genes, loci and alleles.

The point of this exercise was to get you to wrestle with real life data using only the skills you are being shown in this course. What i hope this has illustrated is the surprising amount of information you can extract from data just by looking at its logical structure.

If I do not get murdered by any of you people, i shall re-use this exercise on the next set of victims. So, if i remain alive, i am open to suggestions about how i could set out the coursework better. I haven't done it before so—although i am fairly confident that it's a good idea—i don't delude myself that i have got it absolutely right first time.

Finally i don't want you to lose sight of the fact that the coursework i set you—and which you have all got stuck into most commendably—could test only a small part of the material on offer in the course materials. I am going to assume that by the time you come to the exam you will have attempted all the exercises in the course materials. What else can i do, after all? If that sounds harsh bear two things in mind: (i) you can pester me with questions about them at any time. (ii) the course materials are a lot smaller than the kind of textbook you would otherwise be attempting to swallow. How many people can you fit in a telephone box? No, don't answer that question!

Good luck