Tutorial Sheet 3. Dummy variables

October 5, 2017

The usual (or 'natural') language of common discourse has been developed over millennia to be easy and efficient in a number of ways. Efficiency here includes being able to express *fairly* complex propositions rather succinctly, and also to serve as an aid for a reader or listener to understand the point at hand, for example to prevent long lists of definitions being needed. Natural language is not 'designed' so much as 'evolved' and it has evolved for common day speech, not specifically for a precise scientific application, such as mathematics. Thus natural language allows us to speak of common objects easily but does not easily allow us to talk easily about *new objects no one has ever thought about before.* The latter is what we aim to do in our mathematics, and our language is not always particularly good at it.

Thus we start to see that our mathematics and mathematical notation, at least in an ideal form, is at odds with our usual use of language. To get round this problem, mathematicians increasingly use a special version of English tailored to their own purposes, and some of the differences between ordinary 'natural' English and mathematical English can be surprising for beginners.

One of the ways in which common English obtains a high efficiency is by having fixed names for a large number of fixed objects. Mathematics does the same, *some of the time.* So for example, we are used to always having the letter x represent the 'unknown quantity to be found', π always equal to $3.1415926\ldots$, and i always to be the square root of -1. Therefore if we *always* use these letters in this way we can save some considerable work in communicating our mathematics, relying instead on the listener's understanding of how these letters are used. Of course this assumes some knowledge on behalf of the reader.

However, this isn't always appropriate. In Jones' polynomial for the primes, i is not the square root of -1 and x is not the only unknown and certainly not the most important one. Thus our mathematical language is more powerful and more flexible if we don't *always* have built-in conventions that must always be used. Ultimately mathematics can be hugely powerful in such a language, but this depends on being able to discard popular conventions about π, i and x, and others, *when necessary.*

Thus for example authors, when giving a formula involving e, might be best advised to add 'where e is $2.71828\ldots$' at least once in their document to make the meaning clear, for indeed in many contexts e is the charge on the electron, or the eccentricity of an ellipse, or something else. (However, students answering a question can rely on the context of the question to distinguish what is meant by the various letters.)

The rule that any letter can be introduced, but its meaning must be explained, does not just apply to the teacher or lecturer. It applies to the student too. So going the other way, in an answer to an exercise or assignment a student is allowed to use unusual letters whenever he or she likes. But of course if a new letter is introduced, it must be explained (or even defined). I have come across students who didn't believe it was possible (or 'allowed') to do this. Not only is it allowed, but in some cases it is necessary!

But there is a more serious place where natural language and mathematical language differ greatly, and that is where mathematical language needs a quickly introduced and *ad hoc* name for a new object not considered up to this point. This situation occurs very very often, and most students of mathematics will have seen such examples frequently, often without even realising it. Being aware of such matters will make reading and writing mathematics much more pleasant and straightforward.

One such example that causes immediate problems for many when they reach university is in summation notation, such as

$$\sum_{n=1}^{100} n^2 = 338350.$$

My understanding of pre-university mathematics, at least as taught in practice, is that teachers have noticed that questions at that level (GCSE or A-level exams) are always set in such a way that the letter under the summation sign is always the same. (I have used n here, which is probably the most common choice.) Therefore students are taught to handle summations *with that particular letter only*. These students can (and often do) get into a horrible muddle when that letter is changed. Thus in extreme cases a student familiar with the above equation (and most likely one who would be able to calculate the number given from the information on the left side of the =) finds the question

$$\text{Find the value of } \sum_{k=1}^{100} k^2.$$

outside his/her experience or capability.

In this case, the letter k is one of these letters that has been introduced rapidly as an *ad hoc* name for a new object, namely the *index* of the summation. We imagine a varying quantity *temporarily called* k that varies over the integers between 1 and 100 inclusive and sum up the values k^2. Sometimes such variables k are called *dummy variables* or *bound variables*.

The problem some students seem to have is that, because of the way they have been taught about summations perhaps, they do not realise that in $\sum_{n=1}^{100} n^2$ the letter n is *also* a dummy variable. To be sure, many of these issues are at a subconscious level and so actually articulating them is not always easy but I suspect that many students think about summation as, 'when doing summation you add up the values for all of the possible indices n'. The letter n, to many people has a 'fixed' meaning that always means 'the index of a summation when summations are being done' just like π has a 'fixed' meaning $3.14159265\ldots$. This is a misunderstanding of the mathematical notation, perhaps encouraged by a limited number of examples seen at pre-university level.

To be sure, there are some subtle issues here, but the point is that at university, even at the very beginning of a university course, some flexibility in choice of letters is necessary, and may even be assumed by the lecturer, even when that sort of flexibility is outside the students' experience to that point.

You know when you are faced with a dummy variable or bound variable when changing the name of the variable does not change the meaning in any way. Thus n in $\sum_{n=1}^{100} n^2$ is a dummy variable because $\sum_{n=1}^{100} n^2$ means exactly the same as $\sum_{k=1}^{100} k^2$. Another test for whether a variable is a dummy variable or not is to ask, is it sensible to ask what its value is? Note that it is completely meaningless to ask what the unique value of n in the expression $\sum_{n=1}^{100} n^2$ is. There is none, and in fact to ask this question is the same as asking what the unique value of n in the expression $\sum_{k=1}^{100} k^2$ is.

Another example where dummy variables come up (and cause problems) is in proof by induction. I have seen students who have had as part of their pre-university training some idea of what mathematical induction is and who were capable of writing out perfectly adequate proofs by induction. But the training was so mechanical, given in an entirely algorithmic way such as 'When the question asks you to prove something by induction, write this down then this. . . ' that the student was only able to do induction when the induction letter was k. When asked to prove a statement by induction *on a different letter* the result is a disaster. Of course the objective at university should be to teach mathematics in a way that is understood and adaptable to other situations. This is the 'transferable skills' objective that is perhaps the key objective for any university education, whatever the student's goals or aspirations are.

Another place where you will already have seen dummy variables is in definite integration. So the x in $\int_0^1 x^2\,\mathrm{d}x$ is a dummy variable, and we can see this because changing x to y gives $\int_0^1 y^2\,\mathrm{d}y$ which has the same value, 1/3.

(The cases of differentiation and indefinite integration are much more subtle, because the 'x' in the 'function x^2' for example is already a dummy variable. We will return to this at a later time.)

One conclusion to all this is that mathematical notation should not and usually does not confuse or place a straightjacket on what you can and can't do, but rather should be specifically designed to be extensible and adaptable to different situations. (There is always the proviso that if you do extend notation you must explain the new version of it.) A simple example is modifying the summation notation to allow different letters simultaneously. If this is done new facts such as

$$\sum_{n=1}^{100} \sum_{m=n}^{100} (n+m) = 510050$$

are expressible. In other words, students should look for ways to *generalise* the notation they see, and teachers should point them in these directions. The topic of generalisation or abstraction is a major one in the methodology of mathematics and will be taken up again in many places.

Finally, we introduce a difficulty that arises out of common language and a popular but imprecise version of this used quite frequently in mathematics. That is the question of constants and variables—what these two words mean and how to recognise a quantity as being a 'variable' as opposed to a 'constant'.

It is common to think of some quantities as being 'variable', i.e. changing. But it is not clear what this means as our mathematics itself does not change in time. Thus t might represent 'time' and $v = v(t)$ 'speed', as a 'function' of time. But our mathematics looks at the value of v at a specific moment t, so neither t (usually called the 'independent variable') nor v (the 'dependent variable') actually *change* in the mathematics. If we need to look at the 'rate of change', a strange term for something that is not actually changing, we can do so by using differentiation. The function $v(t)$ is differentiated with respect to t and expressed as some other function, $a = a(t)$ for the acceleration, so that the rate of change of v at a particular moment t is expressed by $a(t)$. Once again this points to something subtle about differentiation (and its inverse operation, indefinite integration) which is something we will have to revisit later.

So much for variables. For constants, consider e. If e is Euler's number, that is surely a constant. But if e is the charge on an electron, is e constant or changing? Recent theories in physics say that to the best of our knowledge the charge on an electron is also constant, but no one is completely sure about this in the way that we are that Euler's number won't change.

My conclusion is that the words 'variable' and 'constant' are there as part of a special kind of efficiency of human language, to help aid *intuition* about a *situation* that in other respects has no normative role on how the letter behaves.[1] Such words may also give *hints* as to what sort of work will be done later with these letters. In other words, these words are optional, but may help the reader considerably when used well, and may become essential if the clarity of the rest of the writing is not up to scratch.

Exercise 1. What is the value of $\sum_{i=1}^{10} \sum_{j=i}^{10} j$?

[1] Something is 'normative' if it says what some other something should do or ought to do or must do. This word is *not* the same as 'normal'.