ind.1 Inductive Definitions

mth:ind:idf:sec In logic we very often define kinds of objects *inductively*, i.e., by specifying rules for what counts as an object of the kind to be defined which explain how to get new objects of that kind from old objects of that kind. For instance, we often define special kinds of sequences of symbols, such as the terms and formulas of a language, by induction. For a simpler example, consider strings of parentheses, such as "(()(" or "()(())". In the second string, the parentheses "balance," in the first one, they don't. The shortest such expression is "()". Actually, the very shortest string of parentheses in which every opening parenthesis has a matching closing parenthesis is "", i.e., the empty sequence $\emptyset$. If we already have a parenthesis expression p, then putting matching parentheses around it makes another balanced parenthesis expression. And if p and p' are two balanced parentheses expressions, writing one after the other, "pp'" is also a balanced parenthesis expression. In fact, any sequence of balanced parentheses can be generated in this way, and we might use these operations to *define* the set of such expressions. This is an *inductive definition*.

Definition ind.1 (Paraexpressions)**.** The set of *parexpressions* is inductively defined as follows:

1. $\emptyset$ is a parexpression.
2. If p is a parexpression, then so is (p).
3. If p and p' are parexpressions $\neq \emptyset$, then so is pp'.
4. Nothing else is a parexpression.

(Note that we have not yet proved that every balanced parenthesis expression is a parexpression, although it is quite clear that every parexpression is a balanced parenthesis expression.)

The key feature of inductive definitions is that if you want to prove something about all parexpressions, the definition tells you which cases you must consider. For instance, if you are told that q is a parexpression, the inductive definition tells you what q can look like: q can be $\emptyset$, it can be (p) for some other parexpression p, or it can be pp' for two parexpressions p and $p' \neq \emptyset$. Because of clause (4), those are all the possibilities.

When proving claims about all of an inductively defined set, the strong form of induction becomes particularly important. For instance, suppose we want to prove that for every parexpression of length n, the number of (in it is $n/2$. This can be seen as a claim about all n: for every n, the number of (in any parexpression of length n is $n/2$.

Proposition ind.2. *For any n, the number of (in a parexpression of length n is $n/2$.*

Proof. To prove this result by (strong) induction, we have to show that the following conditional claim is true:

> If for every $k < n$, any parexpression of length k has $k/2$ ('s, then any parexpression of length n has $n/2$ ('s.

To show this conditional, assume that its antecedent is true, i.e., assume that for any $k < n$, parexpressions of length k contain k ('s. We call this assumption the inductive hypothesis. We want to show the same is true for parexpressions of length n.

So suppose q is a parexpression of length n. Because parexpressions are inductively defined, we have three cases: (1) q is $\emptyset$, (2) q is (p) for some parexpression p, or (3) q is pp' for some parexpressions p and $p' \neq \emptyset$.

1. q is $\emptyset$. Then $n = 0$, and the number of (in q is also 0. Since $0 = 0/2$, the claim holds.

2. q is (p) for some parexpression p. Since q contains two more symbols than p, $\text{len}(p) = n - 2$, in particular, $\text{len}(p) < n$, so the inductive hypothesis applies: the number of (in p is $\text{len}(p)/2$. The number of (in q is $1 +$ the number of (in p, so $= 1 + \text{len}(p)/2$, and since $\text{len}(p) = n - 2$, this gives $1 + (n - 2)/2 = n/2$.

3. q is pp' for some parexpression p and $p' \neq \emptyset$. Since neither p nor $p' = \emptyset$, both $\text{len}(p)$ and $\text{len}(p') < n$. Thus the inductive hypothesis applies in each case: The number of (in p is $\text{len}(p)/2$, and the number of (in p' is $\text{len}(p')/2$. On the other hand, the number of (in q is obviously the sum of the numbers of (in p and p', since $q = pp'$. Hence, the number of (in q is $\text{len}(p)/2 + \text{len}(p')/2 = (\text{len}(p) + \text{len}(p'))/2 = \text{len}(pp')/2 = n/2$.

In each case, we've shown that teh number of (in q is $n/2$ (on the basis of the inductive hypothesis). By strong induction, the proposition follows. □

Photo Credits

Bibliography