## **Induction**

### **1 Introduction**

**1.** *Induction* is the process of justifying quantified, categorical generalizations such as "All dogs like hot dogs." and "92% of Canadian adults are owners of a mobile phone." based on data about particular cases which we have experienced. 

Claims like this are made all the time by people in everyday life. For example, a lot of common knowledge about things are their properties is encoded in generalizations, such as "Birds have wings." and "Bananas grow on trees.".

Stereotypes about different nations or ethnicities, such as "All Irish people love a drink.", are generalizations. So are superstitions such as "I always play well when I wear my lucky socks.".

As the examples of stereotypes and superstitions show, induction in everyday life is often done hastily. Doing it properly requires collecting a large, unbiased sample, so that the percentage discovered in the sample is likely to be close to that in the population. When a categorical proposition has a quantity that is universal or near-universal it can be used in inferences which classify new objects or events. For example, if you know "Almost all dogs have tails.", you can infer that Jack's new dog, Jim, has a tail.

### **2 Inductive Generalization (IG)\***

**1.** Induction is the process of justifying *quantified, categorical* propositions such as "All dogs like hot dogs." and "92% of Canadian adults are owners of a mobile phone." based on information about particular cases which we have experienced.

People make claims like this all the time. For example, a lot of common knowledge about things and their properties is encoded in such propositions, such as "Birds have wings." and "Bananas grow on trees.". Stereotypes about different nations or ethnicities, such as "All Irish people love a drink.", are also quantified categorical propositions.

These propositions are *categorical* in that they are about *categories* or *classes* or *types* of thing, rather than a particular *case* or *instance* of that type. For example, "Jim loves chasing squirrels." is about a specific dog, Jim, while "Most dogs are things that love chasing squirrels." (or more naturally "Most dogs love chasing squirrels.") is about dogs in general.

These proposition are *quantified* in that they specify what proportion or percentage of members of the initial class belong to the class mentioned in the predicate. "Nine out of ten dentists brush with Oral-B toothbrushes." tells us that the percentage of dentists who use an Oral-B toothbrush is 90%. If the quantity is "All" or 100%, or "None" or 0%, the proposition is *universal*. Universality is rarely the case; what we more often get is a proposition describing a probabilistic relation—e.g. if F is present, G is present in 90% of cases. We are happy if the frequency of joint appearance or non-appearance is very high or *near-universal*.

**2.** We turn now to the process of generalization from a sample to a wider population. Consider the following scenario:

Jack shakes a large opaque basket filled with 4,000 black and red cubes, reaches in without looking, and grabs 500. He counts the reds, sees that he has 450, and then on this basis infers that roughly 90% of the cubes in the basket are red.

Jack's inference is an instance of *inductive generalization* (IG) (or sometimes simply *induction*), and in standard form (i.e. with the premises above the line and the conclusion below) it looks like this:

1.    Cube1 … Cube500 are all cubes in the basket.

2.    90% of the 500 cubes examined are red.

J    ----------------------------------------------------

3.    Roughly 90% of the 4,000 cubes in the basket are red.

(*Important Note*: This analysis of the inference does *not* use the literal propositions in the passage. Rather, the relevant information is *extracted* from the passage. Which information is important is about to be explained.)

**3.** This inference concerns a *sample* of cubes (500 of them) from a wider *population* (of 4,000). The population is all the cubes in the basket, and this is mentioned in the conclusion. The sample is the cubes Jack looked at, and this is mentioned in premise (1).

Since we are interested in the percentage of the cubes that are red, the cubes can divided into two types: those that are red, and those that are not red. The color of the cubes is a *variable*, which means that it can take multiple *values*, in this case two: red and black. The information about each cube can be recorded in a table, as follows:

Color

Red    Black

Cube 1    \*

Cube 2    \*

Cube 3            \*

   ...

   ...

Cube 499            \*

Cube 500    \*

Writing out all of the information in this table in propositions would be a lot of work; there would be 500 premises stating that each cube is a cube in the basket (i.e. (1) Cube 1 is a cube in the basket. (2) Cube 2 is a cube in the basket. …) and 500 more stating the color of each cube (i.e. (1) Cube 1 is red. (2) Cube 2 is red. (3) Cube 3 is black. …). What we do instead is summarize all of this information in two premises. Premise (1) states that the 500 cases are cubes in the basket, while premise (2) states the proportion that are red. (The remainder are then assumed to be black.)

**4.** From the fact that 90% of the cubes in the sample are red, Jack infers that roughly 90% of the population of cubes are red. That is, he generalizes. The conclusion moves beyond the specific cubes which were examined to cubes in the basket generally.

Here is the general form of IG:

1.    Case1  … casen are all F.

2.    % of case1  … casen  are also G.

3.    The sample is large.\*

4.    The sampling method yields an unbiased sample.\*

J    ----------------------------------------------------------

5.    Roughly % of cases of F are G.

"F" and "G" stand for any two types of thing; they can refer to *either the presence or the absence* of any type of thing. The first two premises refer to a limited number of cases of F, while the conclusion refers to all Fs. The sample is numbered from 1 to n. The percentage-sign (%) stands for a proportion, expressed as a percentage (or sometimes a fraction, and in ordinary speech by a quantifying word or phrase such as "All", "Most", "A majority of", "Some", and so on). The word "roughly" (or some equivalent word) appears in the conclusion because it is improbable that the percentage of Gs in the population is exactly the same as the percentage of Gs in the sample.

IG can be used whether F and G are described positively or negatively. For example, we might be interested in the percentage of cases in which something is absent that are also cases where a second thing is absent (e.g. In 100% of places where water is absent, life is impossible), or one thing is absent and another present (e.g. 72% of buildings without sprinkler systems suffer serious damage in fires) or the first thing is present and the second absent (e.g. 97% of children who have been vaccination do not contract a certain illness).