Probabilistic characterization of the Sisters' Paradox

by

Jukka Corander[1]

A problem sometimes encountered in quiz books or magazines is so called Sisters' Paradox. In its particular formulation You are told about a family with two children and that they happen to have a daughter, but You don't gain direct knowledge about the gender of the other child. The question then stated is concerned with the probability that also the other child is a girl. I have earlier done an experiment by asking this question from two separate audiences; one consisting of people with rigorous mathematical background, the other one consisting of scientists in general (ranging from economics to history). Without exceptions, members of the mathematical audience answered that the probability equals 1/3 and members of the other audience answered that it equals 1/2. The general audience referred to common sense when motivating their answer, but the mathematically oriented audience backed up their reasoning with a calculation referring to the very definition of conditional probability: $P(C|A) = P(C \cap A)/P(A)$, where A refers to the event of observing one girl and C to the event that there are two girls in the family. Assume families with two children are represented by the following fractions (for more details see below) with respect to the family configuration: $P(BB) = 1/4, P(GG) = 1/4, P(BG) = 1/2$, where '$B$' and '$G$' now refer to a boy and a girl, respectively. When these fractions are plugged into the formula of conditional probability, the the result indeed equals 1/3 (check this!).

Most maths sources present this problem as an example of the difficulties mathematically untrained people have concerning understanding conditional probabilities. In roughly 99 out of 100 sources, the solution given to the problem equals 1/3, and these sources typically claim that people who claim the answer to equal 1/2, are blatantly wrong. In fact, more rarely any sources (one is Dr. Math at the web) provide an analysis of the problem at more depth, and show that both 1/3 and 1/2 are correct answers, and moreover, that 1/2 can be a more reasonable answer under certain circumstances.

The ambiguity arises from the fact that the problem formulation doesn't state exactly in probabilistic terms how You gained the information about the existence of the girl in this particular family. To really solve the problem in a more realistic way, one needs to assume a probabilistic model which generates Your observation. One feasible model is the following. To simplify the problem, assume that boys (B) and girls (G) have equal birth and survival rates in the population where you made the observation, i.e. we expect that the families with two children are represented with the following proportions of configurations: $P(BB) = 1/4, P(GG) = 1/4, P(BG) = 1/2$. Assume now You gather the information through the following mechanism. You sample randomly one apartment in a randomly chosen city in the population. Then You go to the door and ring the doorbell. The door is opened by a girl and you ask whether

[1] The author is indebted to professor Elja Arjas for introducing this controversial issue once upon a time.

she has exactly one sibling, and she answers, Yes. Now You wish to calculate the conditional probability that the sibling is also a girl. The solution is derived by the celebrated Bayes' theorem. Denote by A your observed event (the one girl). The conditional probability is defined as

$$P(GG|A) = \frac{P(A|GG)P(GG)}{P(A|GG)P(GG) + P(A|BG)P(BG) + P(A|BB)P(BB)}, \quad (1)$$

where $P(A|\cdot)$ is typically called the *likelihood* of the observed event, given the unobserved event in the condition, and $P(BB), P(GG), P(BG)$ are the *prior* probabilities for the unobserved event. Let's take a closer look at the likelihoods. Obviously, $P(A|BB) = 0$, because the observation is impossible given this event, and similarly, $P(A|GG) = 1$, as the observation is certain given this event. $P(A|BG)$ is a more tricky one, however, it is a reasonable model to assume that a randomly chosen child in the family comes and opens the door. Then, the conditional probability equals

$$P(GG|A) = \frac{1 \cdot \frac{1}{4}}{1 \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot \frac{1}{4}} = \frac{1}{2}. \quad (2)$$

We can also easily reach the solution 1/3 through this line of reasoning, namely, assume that the families with a boy and a girl have made an agreement upon themselves, such that the girl will always open the door when somebody rings the doorbell. Then, the likelihood $P(A|BG) = 1$, and consequently, $P(GG|A) = 1/3$.

The moral of this story is that mathematical training leads us easily to think about probabilities from the theoretical perspective, i.e. as given quantities to which we apply the laws of probability without much reflecting upon the fact how the information given in a particular problem arose. However, when doing statistical learning, we ought to embrace ourselves with such activities!