In this section of Resonance, *we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. "Classroom" is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.*

M G Nadkarni
Department of Mathematics
University of Mumbai
Mumbai 400 098, India.

The Rule of Three, Maximum Likelihood, and the Second Law of Thermodynamics

A marine biologist wanted to estimate the number of fish in a large lake. He threw a big net and found 3000 fish in the net. He marked all the fish red and released them into the lake. The next morning he again threw a net and this time caught 4000 fish of which 200 were found marked red. What, approximately, is the number of fish in the lake?

This is a well-known example discussed (with different data) in Feller's book [1] on probability as an application of the hypergeometric distribution, and in essentially the above form, it was a question in the selection test for research scholars in physics at a leading research institute in India.

There is considerable compartmentalisation of undergraduate and postgraduate education in science in India. The UGC norms for appointment, as practised at many universities, make it impossible for a doctorate in mathematics with an undergraduate degree in physics and a master's degree in statistics (or any such combination) to obtain employment in our universities and colleges and, thereby, discourages many young minds from widening their knowledge or pursuing their interests. The purpose of this article is to highlight the harmfulness of this situation by discussing the above example.

Indeed a discussion of the example allows us to bring into focus in a very elementary way the unity and interdependence that exists between some very basic elements of mathematics, statistics and physics.

In mathematics one generally pays attention to the most essential facts, idealizes the situation, reducing as many peripherals as possible and arrives at an ideal solution. If we do this to the above example, we can rephrase it as a problem for a grade four pupil.

A lake has certain number of fish of which 3000 are red in colour and others are white. A catch of 4000 fish has 200 red fish in it. What is the total number of fish in the lake?

The question is now categorical, and tentativeness of the previous formulation (in asking for an approximate number of fish in the lake) is dispensed with. A grade four pupil can now answer the question by applying one of the very first mathematical ideas he learns, namely, the rule of three, and can come up with the solution

$$\frac{4000}{200} \times 3000 = 60000$$

for the number of fish in the lake.

How good is this estimate? Can one bet on it? The answer is provided by R A Fisher's maximum likelihood method of estimation, again a very basic idea in statistics taught early in any first course in probability or statistics.

Let N be the number of fish in the lake, an estimate of which we are looking for. Let P_N denote the probability of getting exactly 200 fish with red mark from a draw of 4000 fish from the lake. Now P_N is a function of N and the $N = N_0$ (say) which maximises it is P_N, by definition, Fisher's maximum likelihood estimate of N. When $N = N_0$ the event of getting 200 fish marked red in a catch of 4000 fish is the most likely.

The number of ways a draw of 4000 fish can be made from a total of N fish is

$$\binom{N}{4000}$$

while the number of draws of 4000 fish with exactly 200 fish with red mark is

$$\binom{3000}{200}\binom{N-3000}{3800}$$

Now assuming uniform distribution of fish in the lake i.e., assuming that the probability of a fish being caught in the net is the same for all the fish in the lake, we get the expression

$$P_N \frac{\binom{3000}{200}\binom{N-3000}{3800}}{\binom{N}{4000}}$$

by the usual procedure of taking the ratio of number of favourable events to total number of events. Now P_N increases with N until $N = N_0$ after which it decreases with N. So one considers the 'likelihood ratio'

$$\frac{P_N}{P_{N-1}} = \frac{(N-3000)(N-4000)}{(N-6800)N}$$

which is ≥ 1 or ≤ 1 depending on whether $N \leq N_0$ or $N \geq N_0$. So the largest N for which the likelihood ratio is ≥ 1 is the maximum likelihood estimate of the number of fish in the lake. Now

$$\frac{P_N}{P_N - 1} \geq 1 \Leftrightarrow (N-3000)(N-4000) \geq (N-6800)N\,,$$

and we see after an easy calculation that the largest N for which this inequality is valid is 60000, the same number we got by the rule of three. Thus our mathematical solution agrees with and is justified by the statistical argument of maximum likelihood.

But where is physics in all this? After all the problem was set in a physics test, and a rule-conscious person can file a suit for setting a question from outside the syllabus. Physics enters here in a very fundamental way, for our mathematical solution based on the rule of three and its statistical justification depend on our implicit and innate belief in the second law of thermodynamics in a very broad sense. For, although it is not mentioned, it is understood that at the time of making the first draw from the isolated system of the fish, the lake at its immediate surroundings is in equilibrium and that the fish in the lake are in a state of maximum disorder (what we have called uniform distribution above). The process of making a draw of 3000 fish and releasing them back into the system disturbs this equilibrium, hence the waiting time until the next day (called relaxation time in physics), which is again a proof of our innate belief in the second law of thermodynamics, namely that an isolated system left to itself will eventually be in equilibrium.

Note the interesting role reversal. It is customary to use statistical analysis to confirm or reject a proposed scientific principle, while here an established law of physics is invoked to justify an application of a statistical principle.

So the purpose achieved, we close this article.

Suggested Reading

[1] W Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, Second Edition, Wiley-Eastern, pp. 41–45, 1993.

On Wealth

I am absolutely convinced that no wealth in the world can help humanity forward, even in the hands of the most devoted worker in this cause. The example of great and pure individuals is the only thing that can lead us to noble thoughts and deeds. Money only appeals to selfishness and irresistibly invites abuse.

Can anyone imagine Moses, Jesus, or Gandhi armed with the money-bags of Carnegie?

Albert Einstein