Flat Earth Guide to Probability and Statistics

Prof John C Turner
U.S. Naval Academy
Annapolis, MD
jct@usna.edu

We should include principles from mathematical modeling in the teaching of probability and statistics.

Example: Trajectory of a thrown ball.

Choices: (1) Parabola, (2) portion of a circle, (3) portion of an ellipse.

Parabola fits basic physics and the coefficients are easily related to g, v0 and x0. It also allows the calculation of max height, range and hang time. If our trajectories appear to be asymmetric, we can adjust the formula to allow for wind resistance.

Scope of the model: Can we apply the model in all cases? Not for objects with high v0 because they go into orbit.

What went wrong? We ignored the assumptions behind the parabola, mainly the assumption that the earth is flat. However, for many purposes, the earth is flat. We use maps instead of globes in our cars, for example.

George Box: "All models are wrong. Some models are useful."

Probability distributions are like the parabolic model for the trajectory. They are based on assumptions and must be reasonable models.

Statistics is based on probability models. What are the underlying probability models? Are these models reasonable? Do these models lead to useful conclusions about the original problem? If our probability model is flawed, how does this impact our conclusions?

Impact on teaching:

(1) Assumptions – it is not enough to state the assumptions behind a distribution. It is certainly not enough to say only that when these assumptions are not met, the distribution is not the correct one. The assumptions are never met, so we should not be doing probability at all? We should consider the extent to which the assumptions are met. We should also consider at least the direction of the effect on our results caused by failure to satisfy the assumptions.
(2) Estimation – The parabolic model depends on g, v0 and x0. These will have to be measured (estimated). What is the effect on our results of this estimation? What is easy to estimate and what is hard to estimate?
(3) Why did we do this? – The models – both probability and statistics – are constructed for a purpose. How does the model help answer the original problem?