Why the Simplest or the Most Beautiful Solution Is Often the Best

Vanessa Chavez de la Rosa, Zaul Gomez, and Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA
vcchavezde@miners.utep.edu, zgomez4@miners.utep.edu, vladik@utep.edu

Formulation of the problem. How do physicists come up with equation that describe nature – i.e., that provide the best fit for observations? Some of them look for the simplest of possible equations, some look for the most beautiful equations – and somehow all of them come up with exactly the equations that bet fit the data.

These are different criteria. In general, what is simpler is not necessarily more beautiful, and vice versa. However, many different optimality criterion do lead to the exact same result. In this talk, we provide a possible explanation for this rather mysterious fact.

Symmetry: general idea. Why can we make predictions about the world? Because many situations are similar. If we encounter a new situation which is similar to the one we experienced earlier, it is reasonable to predict that the outcome of the new situation will be similar. In physics, this similarity is formalized as *symmetry* – when changing the situation in a certain way does not change the outcome.

A simple example. Numerical value of a physical quantity depends on the choice of a measuring unit. If we use cm instead of m, numerical values change but the quantities remain the same. In general, if we replace the original unit with a c times smaller one, all numerical values are multiplied by c: $x \mapsto T_c(x) = c \cdot x$.

For each dependence $y = f(x)$, the numerical value of y can also change to $C \cdot y$ for some C. To eliminate dependence on the y-unit, we can consider the whole family $\{C \cdot f(x)\}_C$.

What we mean by the best. The best – optimal – means that we have a way to compare two families, and we select the family $a_{\rm opt}$ which is better than all others: $\forall a\,(a_{\rm opt} \succeq a)$.

We should consider final optimality criteria. If several families are the best in this sense, we can use this non-uniqueness to optimize something else. So, in the final optimality criterion, there is only one optimal family.

This leads to an explanation. It is reasonable to assume that the relative quality should be invariant under scaling transformation $T_c : \{C \cdot f(x)\}_C \mapsto \{C \cdot f(c \cdot x)\}_C$: if $a \succeq b$ then $T_c(a) \succeq T_c(b)$. Our main result is that for *any* final scale-invariant optimality criterion $\succeq$, the optimal family is the one which is invariant under scaling.

Indeed, since $a_{\rm opt}$ is the best, we have $a_{\rm opt} \succeq T_{1/c}(a)$ for all a. Thus, since $\succeq$ is scale-invariant, we get $T_c(a_{\rm opt}) \succeq a$ for all a. This means that the family $T_c(a_{\rm opt})$ is also optimal. However, since the criterion $\succeq$ is final, there is only one final family. So indeed, $T_c\,(a_{\rm opt}) = a_{\rm opt}$. For scale-invariance, this means that $a = \{C \cdot x^a\}$ for some a. So, no matter what optimality criterion we use, as long as this criterion is scale-invariant, we get the same class of optimal functions.

We showed it on the example of scaling, but this argument works for all possible symmetries – and this explains why optimizing different criteria often leads to the same solution.