Confirmation Bias in Systems Engineering: A Pedagogical Example

Griselda Acosta[1], Eric Smith[2], and Vladik Kreinovich[3]
[1]Department of Electrical and Computer Engineering
[2]Department of Industrial, Manufacturing, and Systems Engineering
[3]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
gvacosta@miners.utep.edu, esmith2@utep.edu, vladik@utep.edu

Abstract

One of the biases potentially affecting systems engineers is the confirmation bias, when instead of selecting the best hypothesis based on the data, people stick to the previously-selected hypothesis until it is disproved. In this paper, on a simple example, we show how important is to take care of this bias: namely, that because of this bias, we need twice as many experiments to switch to a better hypothesis.

1 Formulation of the Problem

Confirmation bias. It is known that our intuitive reasoning shows a lot of unexpected biases; see, e.g., [2]. One of such biases is a *confirmation bias*, when, instead of selecting the best hypothesis based on the data, people stick to the previously-selected hypothesis until it is disproved. This bias is ubiquitous in systems engineering; see, e.g., [1, 4, 5, 8].

How important is it to take the confirmation bias into account? Taking care of the confirmation bias requires some extra effort; see, e.g., [3, 7, 8, 9] and references therein. A natural question is: is the resulting improvement worth this extra effort? How better the result will we get?

In this paper, on a simple example, we show that the result is drastically better: namely, that if we properly take this bias into account, then we will need half as many experiments to switch to a more adequate hypothesis.

2 Analysis of the Problem

Description of the simple example. Let us consider the simplest possible case when we have a parameter a that may be 0 and may be non-zero, and we directly observe this parameter. We will also make the usual assumption that the observation inaccuracy is normally distributed, with 0 mean and known standard deviation σ.

In this case, what we observe are the values $x_1, \ldots, x_n$ which are related to the actual (unknown) value a by a relation $x_i = a + \varepsilon_i$ ($i = 1, \ldots, n$), where ε_i are independent normally distributed random variables with 0 means and standard deviation σ.

Two approaches. In the ideal approach, we select one of the two models – the null-hypothesis $a = 0$ or the alternative hypothesis $a \neq 0$ – by using the usual Akaike Information Criterion (AIC); see, e.g., [6].

In the confirmation-bias approach, we estimate the value a based on the observations $x_1, \ldots, x_n$, and we select the alternative hypothesis only if the resulting estimate is statistically significantly different from 0 – i.e., e.g., that the 95% confidence interval for the value a does not contain 0.

What if we use AIC. In the AIC, we select a model for which the difference $\text{AIC} \stackrel{\text{def}}{=} 2k - 2\ln\left(\widehat{L}\right)$ is the smallest, where k is the number of parameters in a model and $\widehat{L}$ is the largest value of the likelihood function L corresponding to this model.

The null-model $a = 0$ has no parameters at all, so for this model, we have $k = 0$. For n independent measurement results, the likelihood function is equal to the product of the values

$$\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{x_i^2}{2\sigma^2}\right)$$

of the Gaussian probability density function corresponding to these measurement results x_i. Thus,

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{x_i^2}{2\sigma^2}\right)$$

and so, for this model,

$$\text{AIC}_0 = -2\ln(L) = 2n \cdot \ln\left(\sqrt{2\pi} \cdot \sigma\right) + \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} x_i^2.$$

We assume that $x_i = a + \varepsilon_i$, where the mean value of ε_i is 0 and the standard deviation is σ. Thus, the expected value of x_i^2 is equal to $a^2 + \sigma^2$. For large values n, due to the Law of Large Numbers (see, e.g., [6]), the average $\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2$ is approximately equal to the expected value $E[x_i^2] = a^2 + \sigma^2$. Thus,

$\sum_{i=1}^{n} x_i^2 \approx n \cdot (a^2 + \sigma^2)$ and hence,

$$\mathrm{AIC}_0 = 2n \cdot \ln\left(\sqrt{2\pi} \cdot \sigma\right) + \frac{1}{\sigma^2} \cdot n \cdot (a^2 + \sigma^2). \tag{1}$$

The alternative model $a \neq 0$ has one parameter a, so here $k = 1$. The corresponding likelihood function is then equal to

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x_i - \widehat{a})^2}{2\sigma^2}\right).$$

We select the parameter a that maximizes the value of this likelihood function. Maximal likelihood is the usual way of estimating the parameters, which in this case leads to $\widehat{a} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$. For large n, this estimate is close to the actual value a, so we have

$$\widehat{L} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x_i - a)^2}{2\sigma^2}\right).$$

For this model, $x_i - a = \varepsilon_i$, thus,

$$\mathrm{AIC}_1 = 2 - 2\ln\left(\widehat{L}\right) = 2 + 2n \cdot \ln\left(\sqrt{2\pi} \cdot \sigma\right) + \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \varepsilon_i^2.$$

For large n, we have $\sum_{i=1}^{n} \varepsilon_i^2 \approx n \cdot \sigma^2$, hence

$$\mathrm{AIC}_1 = 2 + 2n \cdot \ln\left(\sqrt{2\pi} \cdot \sigma\right) + \frac{1}{\sigma^2} \cdot n \cdot \sigma^2. \tag{2}$$

The second model is preferable if $\mathrm{AIC}_1 < \mathrm{AIC}_0$. By deleting common terms in these two values AIC_i, we conclude that the desired inequality reduces to $2 < \frac{n \cdot a^2}{\sigma^2}$, i.e., equivalently, to

$$n > \frac{2\sigma^2}{a^2}. \tag{3}$$

What if we use a confirmation-bias approach. In the confirmation-bias approach, we estimate a – and we have already mentioned that the optimal estimate is $a = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$. It is known (see, e.g., [6]) that the standard deviation of this estimate is equal to $\sigma_e = \frac{\sigma}{\sqrt{n}}$. Thus, the corresponding 95% confidence interval has the form $[a - 2\sigma_e, a + 2\sigma_e]$. The condition that this interval does not contain 0 is equivalent to $|a| > 2\sigma_e$, i.e., equivalently, to $a^2 > 4\sigma_e^2$. Substituting the above expression for σ_e into this inequality, we conclude that $a^2 > 4 \cdot \dfrac{\sigma^2}{n}$, i.e., equivalently, that

$$n > \frac{4\sigma^2}{a^2}. \tag{4}$$

Conclusion. By comparing the expressions (3) and (4) corresponding to the two approaches, we can indeed see that the confirmation-bias approach requires twice as many measurements than the approach in which we select the best model based on the data.

Thus indeed, avoiding confirmation bias can lead to a drastic improvement in our estimates and thus, in our decisions.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] J. Bohlman and A. T. Bahill, "Examples of mental mistakes made by systems engineers while creating tradeoff studies", *Studies in Engineering and Technology*, 2014, Vol. 1, No. 1, pp. 22–43.

[2] D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus, and Giroux, New York, 2011.

[3] R. M. Koriat, S. Lichtenstein, and B. Fischhoff, "Reasons for confidence", *Journal of Experimental Psycholology: Human Learning and Memory*, 1980, Vol. 6, pp. 107–118.

[4] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises", *Review of General Psychology*, 1998, Vol. 2, pp. 175–220.

[5] H. Petroski, "Framing hypotheses: A cautionary tale", *American Scientist*, 2003, Vol. 91, No. 1, pp. 18–22.

[6] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[7] P. Slovic and B. Fischhoff, "On the psychology of experimental surprises", *Journal of Experimental Psychology: Human Perceptions and Performance*, 1977, Vol. 3, pp. 544–551.

[8] E. D. Smith and A. T. Bahill, "Attribute substitution in systems engineering", *Systems Engineering*, 2010, Vol. 13, No. 2, pp. 130–148.

[9] W. A. Wickelgren, *How to Solve Problems: Elements of a Theory of Problems and Problem Solving*, Freeman, San Francisco, 1974.