Base Rate Bias and False Positives: Vetting Provocative Radio Transients

Thomas W. Hair, Ph.D. , Florida Gulf Coast University, Fort Myers, Florida

Most searches for alien radio transmission have focused on finding omni-directional or purposefully earth-directed beams of enduring duration. However, most of the interesting signals so far detected have been transient and non-repeatable in nature. These signals could very well be the first data points in an ever-growing data base of such signals used to construct a probabilistic argument for the existence of extraterrestrial intelligence. This poster explains the effect base rate bias could have on deciding which signals to include in such an archive based upon the likely assumption that our ability to discern natural from artificial signals will be less than perfect.

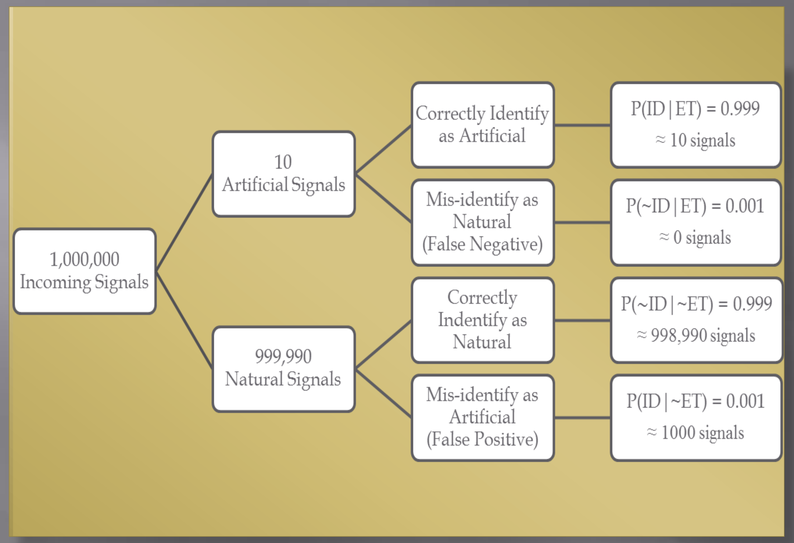
As a thought experiment let us suppose, for example, in a dedicated, wide-spectrum search of the galactic center where 90% of the galaxy's stars lie within just 9% of the sky from our perspective, a one year *long stare* search endeavor yields 1 million radio transient signals of interest. If, unknown to us, there are 10 actual artificial (i.e., extraterrestrial) signals and 999,990 natural signals, then the base rate probability of one random signal from the search being artificial is thus 0.00001 and the base rate probability of a random signal being a natural source is 0.99999. In an attempt to find these artificial signals within the much greater natural ones a complex algorithm must *recognize* their artificialness. In this example, we can imagine an algorithm that has two failure rates of 0.001:

If the algorithm recognizes an artificial signal, it will correctly identify it with a probability of 0.999, and mistakenly fail to recognize it with probability 0.001 (in other words, the false-negative rate is 0.1%).

If the algorithm recognizes a natural signal, it will correctly identify it with a probability of 0.999, but it will mistakenly misidentify the signal as artificial with a probability of 0.001 (the false-positive rate is 0.1%).

So, the failure rate of the algorithmic based transient detection system is always 0.1% in this example.

Now suppose a radio transient is received and recognized by the algorithm as being of intelligent extraterrestrial origin. What is the chance it actually is artificial? Someone exhibiting base rate bias would incorrectly claim that there is a 99.9% chance that this signal is extra-terrestrial and pop the champagne, because the failure rate of the algorithm is always a measly 0.1%. Although this seems to make sense, it is actually faulty reasoning. The application of Bayes' Theorem below shows that the chance the signal is extraterrestrial is actually near 1%, not near 99.9%.



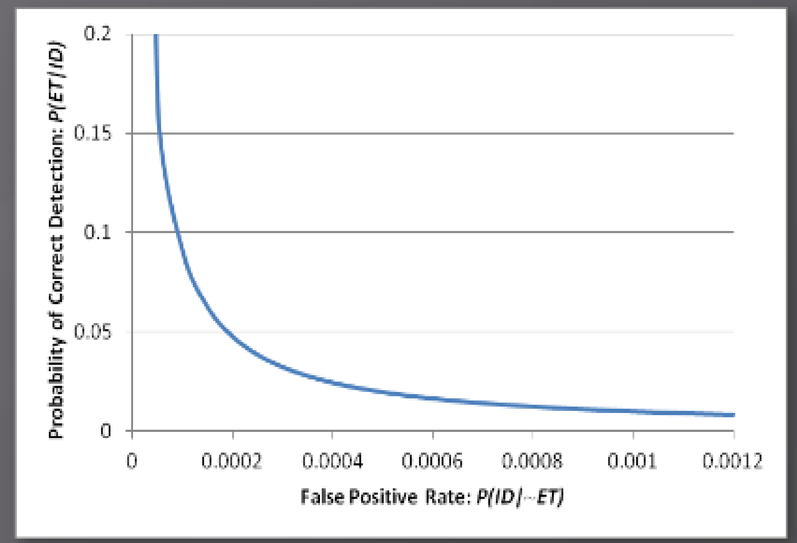
Tabular representation of the base rate bias problem: the algorithm will detect all the artificial transients, but because it also has a false positive rate of 0.001 it will also incorrectly identify 1000 natural signals as artificial. Thus, the actual probability a transient identified as artificial, is indeed artificial, is about 0.01.

Let *P(ET|ID)* be the probability of correctly identifying a transient signal as being artificial, that is, the signal is actually artificial given the algorithm identifies it as such and where *P(ET)* is the base rate probability that any given transient received is artificial, *P(ID)* is the probability a transient is determined by the algorithm to be artificial, and *P(ID|ET)* is the probability the algorithm identifies a transient as artificial given that it is indeed artificial. Therefore, using Bayes' Theorem

$$P(ET|ID) = \frac{P(ID|ET) \times P(ET)}{P(ID)}$$

$$= \frac{P(ID|ET) \times P(ET)}{P(ID|ET) \times P(ET) + P(ID|\sim ET) \times P(\sim ET)}$$

$$= \frac{0.999 \times 0.00001}{(0.999 \times 0.00001) + (0.001 \times 0.99999)} = 0.0099 \approx 1\%$$



As SETI techniques begin to turn away from the search for purposefully Earth-directed signals of enduring duration and more toward finding the ephemeral leakage associated with the complex business of maintaining an interstellar civilization, large or small, then we must come to terms with the fact that most if not all of our detections will be one of a kind…never to repeat with exactly the same parametric characteristics or in the same position in the sky. With this in mind it becomes particularly important to collect a large data set of provocative (i.e., possibly indicative of extraterrestrial intelligence) radio transients through a *long stare* strategy in order to build up a compelling case that they cannot all be explained by natural phenomena in much the same way the dedicated scientists at CERN used their diligence and perseverance to recently claim the Higgs Boson exists to an extremely high degree of confidence. In consideration of this it is important to realize that the finest mathematical sieve will always gather up far more straw from the haystack than it does those very special needles.

Taken in part from:

Provocative Radio Transients and Base Rate Bias: Bayesian Argument for Conservatism

Acta Astronautica 91 (2013) 194-199 , Elsevier Press