Why People Overestimate Small Probabilities?

David Amparan and Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
daamparan@miners.utep.edu, vladik@utep.edu

Formulation of the problem. It is known that people routinely overestimate small probabilities when making decisions. They overestimate the probability of rare side effects – and thus, refuse to take important vaccinations. Experiments performed by the Nobelist Daniel Kahneman and his team show that indeed, most people overestimate small probabilities.

This is a fact, but how can we explain this fact from the biological viewpoint? At first glance, the more adequately we understand the situation, the more adequate decision we can make. So why did evolution preserve this clearly biased perception of small probabilities?

Our explanation. Probabilities are estimates based on our experience. If we saw some event n times out of N, then we estimate the probability as the ratio n/N. But of course, this is only an approximate estimate. If we flip a perfectly symmetric coin 10 times, we may get $n = 5$ heads, but we may also get 6 or 4 or 7.

Which values are possible? If the actual probability is p, then out of N tries, the event happens on average in $\mu \stackrel{\text{def}}{=} p \cdot N$ times, and the variance of number of events is equal to $\sigma^2 = N \cdot p \cdot (1 - p)$. For small p, we have $1 - p \approx 1$, so $\sigma^2 \approx \mu$ and thus, $\mu \approx \sigma^2$. Usually, if we have a distribution with a known mean and standard deviation, we conclude – with high confidence – that the actual value is somewhere between $\mu - k \cdot \sigma$ and $\mu + k \cdot \sigma$, where $k = 2, 3, 6, \ldots$ depending on the desired level of confidence.

So, based on the fact that we observed that the event occurred n time out of N, the only information that we can conclude about the probability p is that $\mu - k \cdot \sigma \le n \le \mu + k \cdot \sigma$, i.e., equivalently, $\sigma^2 - k \cdot \sigma \le n \le \sigma^2 + k \cdot \sigma$, where $p = \sigma^2/N$. By using the known properties of quadratic equations and inequalities, we conclude that

$$\frac{\sqrt{k^2 + 4n} - k}{2} \le \sigma \le \frac{\sqrt{k^2 + 4n} + k}{2}, \text{ so}$$

$$\underline{p} \stackrel{\text{def}}{=} \frac{2n + k^2 - k \cdot \sqrt{k^2 + 4n}}{2N} \le p \le \overline{p} \stackrel{\text{def}}{=} \frac{2n + k^2 + k \cdot \sqrt{k^2 + 4n}}{2N}.$$

We have no reason to consider one of the values from the interval $[\underline{p}, \overline{p}]$ as more probable. So, it makes sense to consider all these values equally possible. In this case, a natural idea is to select the average of these values, i.e., the midpoint $(\underline{p} + \overline{p})/2 = n/N + k^2/(2N)$.

This value is always larger than the frequency n/N – which is the usual (and unbiased) estimate of the actual probability. This provides a possible explanation of why we, in general, overestimate the values of small probabilities.