Klaus Hoechsmann, February 1996

The Likelihood of Unlikely Events.

Suppose your chance of winning the village lottery is one in a thousand. So if you play a thousand times, you can be sure to win, right? "Wrong," you say,"there is always a chance that, by some fluke, I just don't win in this particular run of draws". But you concede that this fluke chance, call it $f(1000)$, is fairly small.

The same fluke factor can come up in many situations. If every thousandth adult in your country is a professor, it would seem reasonable to expect to find such a scholar in any random sample of one thousand citizens. Reasonable yes, but not sure: there may not be any — again with probability $f(1000)$ — do you see? Let us up the ante: if one in ten thousand fellow citizens suffers from a certain affliction, there is neverless a chance, call it $f(10,000)$, of finding none of the stricken in a given poll of ten thousand. How different are $f(1000)$ and $f(10,000)$? Which one is bigger?

Imagine the following generic scenario: the chance of "success" in a certain kind of experiment (e.g., playing a lottery once, testing one specimen from a population) is 1 in n. Then the chance of "failure" in any given trial is surely $1 - 1/n$. Two successive failures would occur with probability $(1 - 1/n) \times (1 - 1/n)$ — do you see? — three of them with probablity $(1-1/n) \times (1-1/n) \times (1-1/n) = (1-1/n)^3$, and so on. The chance of having n independent misses is therefore simply

$$f(n) = (1 - 1/n)^n.$$

It is easy to work out a few examples. Here is a list of some values of n in the left column, with matching fluke factors $f(n)$ on the right.

100	0.366032341	0.37
1000	0.367695425	0.368
10,000	0.367861046	0.3679
100,000	0.367877602	0.36788
1000,000	0.367879257	0.367879
10,000,000	0.367879423	0.3678794
100,000,000	0.367879439	0.36787944
1000,000,000	0.367879441	0.367879441

Surprise, surprise! The first surprise is that all these numbers $f(n)$ are roughly the same. They do creep upward a little, but according to the table above, the creeping from $f(10^r)$ to $f(10^{r+1})$ happens beyond the r-th place after the decimal point. At that rate, these values will never advance as far as 0.367879442. For practical purposes, we might as well think of them as approximating an ideal "limit number" f, lying somehere between 0.367879441 and 0.367879442.

The second surprise is that f is relatively large — about 37%. In other words, you only get a 63% chance of winning, even if you try your luck n times when the odds are 1 to n. This can be discouraging, as n is in the millions for most lotteries. Since $f^{0.7}$ is roughly 1/2, you would need to play $0.7n$ times just to have a 50-50 chance.