Getting hit by lightning ... in the long run

John Kerl

October 10, 2010

Suppose something might happen every day (or week, or month, etc.) with very small likelihood — you getting hit by a bus, a bit randomly flipping from a zero to a one (or vice versa) somewhere on your hard drive, winning the lottery, etc. On each day, you realistically have nothing to worry about. But after a month, or a year, it might eventually catch up with you. Or will it?

1 How do my odds accumulate?

To be precise, suppose the event occurs with a probability p. I'm interested mainly in the case of very small p, since if the event happens with high likelihood each day, then — as we'll see — it'll happen almost certainly in the long run, and that's less worrisome to think about.

Number the days (or weeks, or months, etc.) with $i = 1$ up to n. Let the event occurring on the ith day be A_i, so $P(A_i) = p$. The event occurring on *at least one* of those n days is the "or", or union, of each of the individual events: $\cup_{i=1}^n A_i$. Then from elementary probability, assuming each event is independent of the other (e.g. doesn't influence the next), and a piece of logic called de Morgan's law, we have

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= 1 - P(\sim \cup_i A_i) \\ &= 1 - P(\cap_{i=1}^n \sim A_i) \\ &= 1 - \prod_{i=1}^n P(\sim A_i) \\ &= 1 - (1-p)^n. \end{aligned}$$

So, if something has a one-in-a-million chance (i.e. $p = 1/1,000,000$) of happening each day, then the chance that it'll happen at least once over the course of a year is

$$1 - (1-p)^n = 1 - (1 - 0.000001)^{365} \approx 0.0003649 \approx 1/2739.$$

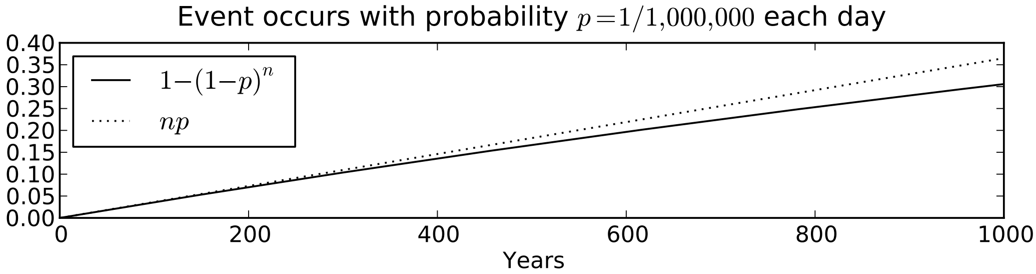
Those one-in-a-million daily odds turned into about one in 2700 for the year[1].

If you use the binomial theorem to expand $(1-p)^n$, then if p is small, $1-(1-p)^n$ is roughly np — although this rough estimate gets less and less correct as either p or n gets larger. (That is, this approximation makes sense when p is small and np is also small.) E.g. here $0.0003649 \approx 365/1,000,000$.

Here's a graphical depiction. For a one-in-a-million daily event, here's what the true happens-in-n-days probability, $1-(1-p)^n$, and the approximation np, look like. Note, though, that I labeled the horizontal

[1] Side note: if p is really really small, then your calculator may round $1-p$ up to 1. Then $1-(1-p)^n$ will come out to be zero.

axis in years — so 1000 years really means 365,000 days. Here it's clear that the np approxmation is not too far off, as long as np remains small.



From the graph you can guess that even for very small p, as n gets bigger (i.e. as more time goes by), $1-(1-p)^n$ eventually gets close to 1, which is to say, close to 100% certainty. If the to-be-avoided event happens with one-in-a-million odds on each day, then as we saw the chances of it happening after a year are 0.03649%; by 80 years, that goes up to 2.9%. But for something that happens with one-in-a-million odds *every second* — there are $n = 60 \cdot 60 \cdot 24 \cdot 365$, or over 31 million, seconds in a year. And then $1-(1-p)^n$ gives you 99.999999% certainty that that thing will have happened at least once by the end of the year.

2 OK, so how long should I wait?

The reasonable next question is, how long will it take for this rare thing to become a likely thing in the long run? For example, how many weeks of buying one lottery ticket a week will it take for me to have a 50-50 chance of having won at least once?

Call your target level of likelihood L — e.g. $L = 0.5$ for 50-50. We want to find n such that $1-(1-p)^n \geq L$: visually, how long it takes the curve in the above graph to reach a particular level, such as 0.5. Solving for n, with logarithms to help out, we have

$$\begin{aligned} 1-(1-p)^n &\geq L \\ 1-L &\geq (1-p)^n \\ \log(1-L) &\geq n\log(1-p) \\ n &\geq \frac{\log(1-L)}{\log(1-p)}. \end{aligned}$$

(It doesn't matter if you use natural logs, or log-base-10, or log-base-2, or whatever, since the scaling constants that relate them to one another cancel out in the fraction. Also note that the direction of the inequality changed between the 3rd and 4th lines since we divided by $\log(1-p)$ which is negative.)

For example, the multi-state Powerball lottery has weekly chance of winning the jackpot of $p = 1/195,249,054$. For $L = 0.5$ (my desired 50-50 level), we get

$$\begin{aligned} n &\geq \frac{\log(1-L)}{\log(1-p)} \\ &\approx 135,336,330 \text{ weeks} \\ &\approx 2,602,621 \text{ years.} \end{aligned}$$

3 So ...

The short story here is that if the odds of something happening each day are 1 in M (by which I mean $p = 1/M$), then don't start worrying until somewhere around $M/2$ days, or maybe $M/10$ to be a more cautious. If you want more confidence, then do the math above to figure your precise odds. And if you want absolute confidence ... well, unfortunately we don't get that. One lightning strike is enough. We can simply take reasonable precautions (like avoiding high places when thunderclouds are near) and enjoy all the good things which *are* quite likely to occur — like enjoying a nice stroll in that autumn rain.