

Re: Please help me to find a mistake here

Re: Please help me to find a mistake here

Source: <http://sci.tech-archive.net/Archive/sci.math/2006-07/msg00999.html>

- *From:* matt271829-news@xxxxxxxxxxxxx
 - *Date:* 6 Jul 2006 10:37:43 -0700
-

valery wrote:

matt271829-news@xxxxxxxxxxxxx wrote:

Anyway, FWIW, making some simplifying assumptions (that seem entirely reasonable since 71 is so much smaller than 2^{32}), I make the probability of event E equal to about $1.76E-10$. I think this is slightly different from James Waldby's answer. I did it in a slightly more elaborate way, but with these particular numbers I would have thought the answers should differ by only an imperceptible amount. I'm not sure at the moment why the answers differ as much as they do, but if I get a following wind I might try to pick through it and figure it out.

$1.76E-10$ sounds just perfect! it passes reality check and is very well aligned with my experiment (taking your numbers gives us about 0.58 probability that event E occurs at least once per 2^{32} attempts). Could you post your formulas, please? (with some explanations if possible ;-).

-Valery.

<http://www.harper.no/valery>

How do you get 0.58? I make $1 - (1 - 1.76E-10)^{(2^{32})}$ equal to about 0.53... Incidentally, I don't think that your eight-out-of-twelve experimental result is particularly diagnostic when discriminating between the various suggested answers. With such a small number of trials the experimental results are consistent with a fairly wide range of actual probabilities.

Anyway. As understand it, to calculate $\Pr(E)$, the probability of event E, we need to select 71 different numbers randomly from the numbers 0 to $2^{32} - 1$, and find the probability that at least two pairs differ by exactly one bit. I will let $k = 71$ and $n = 32$. The exact answer looks rather difficult to compute (unless I'm missing something obvious), but as a simplification one can assume that all the pairs in the set of k numbers can be considered independently.

Re: Please help me to find a mistake here

Re: Please help me to find a mistake here

The simplest way to do this is something along the lines of the method that I think James Waldby used, and I'll use some of his notation for consistency. The probability that any individual pair of (distinct) numbers does NOT differ by exactly one bit is, as far as I can see, given by $r = (2^n - (n + 1)) / (2^n - 1)$. The first number can be chosen anyhow, and the second number can be chosen as any of $2^n - (n + 1)$ out of the remaining $2^n - 1$. However, this is different from James' answer, which I think has $r = (2^n - n + 1) / (2^n - 1)$.

There are $w = k * (k - 1) / 2$ pairs of numbers to consider, and, treating all these as independent, the probability, $R(x)$, that exactly x of these pairs differ by just one bit is given by the binomial distribution, $R(x) = C(w, x) * r^x * (1 - r)^{w - x}$. This gives

$$R(0) = r^w$$

$$R(1) = w * r^{w - 1} * (1 - r)$$

and then

$$\Pr(E) = 1 - R(0) - R(1)$$

Using this method I get $\Pr(E) = 1.7133E-10$, approx. (The difference between this answer and your previous value of $1.5057E-10$ is due to my and James' slightly differing formulas for r .)

You can easily see by experimenting with some small values of n and k that this answer is not exactly right. This is because the pair probabilities aren't really independent as we assumed they were. One would expect the difference to be very small though, since k is so much smaller than 2^n .

I also tried a more elaborate method, as follows:

$$R(0) = 2^n * (2^n - (n + 1)) * (2^n - 2 * (n + 1)) * \dots * (2^n - (k - 1) * (n + 1)) / (k! * C(2^n, k))$$

$$R(1) = 2^n * n * (2^n - 2 * n) * (2^n - 2 * n - (n + 1)) * (2^n - 2 * n - 2 * (n + 1)) * \dots * (2^n - 2 * n - (k - 3) * (n + 1)) / (2 * (k - 2)! * C(2^n, k))$$

Here I am looking at the ways that each of the k numbers can be chosen in turn from the 2^n candidates so as to meet the criteria. So, for $R(0)$, the first number can be any of the 2^n , the second any of $2^n - (n + 1)$, the third any of $2^n - 2 * (n + 1)$, and so on. Then we divide by $k!$ to compensate for multiple counting, and finally divide by the total number of ways of choosing k from 2^n to get the probability. $R(1)$ is calculated in a similar way.

This method is not exact either, because when we reduce the count of candidate numbers by $n + 1$ each time we are potentially double-counting. Using this second method I get $\Pr(E) = 1.7611E-10$, approx. Even though the two answers are closer than before I am still slightly surprised that the difference is this large. Initially I managed to convince

Re: Please help me to find a mistake here

Re: Please help me to find a mistake here

myself that the second method ought to be more accurate, but I'm not sure.

One final point. In your original post you didn't specify whether the 2^{32} random mappings are all distinct. In other words, are they chosen randomly from the total population of permutations with or without replacement? Strictly speaking, the $1 - (1 - \Pr(E))^{(2^{32})}$ formula is correct only if they are chosen with replacement (duplicates possible), but I think the difference here will be truly minuscule for the numbers in question.