Expected amount of repeats in a random sequence of integers

I'm looking at a series of random integers generated by a CSPRNG and noticed that there are more repeats (that is a number is in the sequence 2 or more times e.g. 9,3,8,5,6,3 - 3 is a repeat) than I expected.

I generated 10,000 numbers, each between 1 and 100,000, this resulted in 9,516 unique numbers. Does this seem correct, and if so, how would I calculated the expected about of unique numbers for n random numbers of a range 1 to x?

(probability) (sequences-and-series) (random)

edited Aug 6 '15 at 12:12

asked Aug 6 '15 at 12:07
Paul
13 3

As Bernhard says in his answer, this does seem about correct. Random numbers are counter-intuitive in their behaviour. They "cluster" more than we tend to expect. – Colm Bhandal Aug 6 '15 at 12:38

This question is a refinement of the Coupon Collector problem, which has been much studied. The usual problem asks how many trials you'll need before you can expect to see all the numbers, but of course you can also ask for the expected number of coupons you'll see after n trials, number of multiples, and so on. Easy to get references on line; here, for example, is one: math.ucla.edu/~pak/courses/pg/l10.pdf – lulu Aug 6 '15 at 13:05

3 Answers

When sample n times from the set $\{1, \ldots, x\}$, then the expected number of unique values is $x[1-(1-1/x)^n]$. With $n=10000$ and $x=100000$, this gives approximately 9516.303.

answered Aug 6 '15 at 13:16
Byron Schmuland
29.7k 3 59 133

Thank you - I've run several sequences with different x and n values and put the same into your formula - very close match each time! – Paul Aug 6 '15 at 15:10

The probability a given single number does not appear at least once is
$\left(1-\frac{1}{100000}\right)^{10000} \approx e^{-1/10}$

so the expected number not appearing is 100000 times this, near 90483.7,

making the expected number of unique numbers about 9516.3, surprising close to what you observed.

answered Aug 6 '15 at 13:16
Henry
73.1k 3 44 107

Thank you - I accepted the other answer just because I understood the formula, but I can see you have the same answer. Tried upvoting you but could not as new here. – Paul Aug 6 '15 at 15:12

Your question is strongly related to the birthday paradox. Calculating the expected value of unique numbers could be done, e.g., by using combinatorial arguments. But as to your question: Yes, such a number indeed seems to be plausible.

answered Aug 6 '15 at 12:18
Bernhard
807 2 5

Thank you for the link – Paul Aug 6 '15 at 15:13