Great Expectations, or: Expect More, Work Less

Steven J Miller
Williams College

Steven.J.Miller@williams.edu
http://www.williams.edu/Mathematics/sjmiller/

Mount Holyoke College, October 13, 2011

Clicker Questions

Collect all four!

Cereal toy problem

A cereal company decides to put one of N toys in each specially marked box. Each box has exactly one toy, and each box is equally likely to have any of the N toys. For N large, approximately how many boxes do we expect to buy before we have one of each toy?

Collect all four!

Cereal toy problem

A cereal company decides to put one of N toys in each specially marked box. Each box has exactly one toy, and each box is equally likely to have any of the N toys. For N large, approximately how many boxes do we expect to buy before we have one of each toy?

- (a) Around 6.
- (b) Around N.
- (c) Around $N \log N$.
- (d) Around $N^{3/2}$.
- (e) Around $N^2/\log N$.
- (f) Around N^2.
- (g) More than N^3.

Prime divisors

Number of prime divisors

Let N be a large number. If we choose an integer of size approximately N, on average about how many distinct prime factors do we expect N to have (as $N \to \infty$)?
It might be useful to recall the Prime Number Theorem: The number of primes at most x is about $x/\log x$.

Prime divisors

Number of prime divisors

Let N be a large number. If we choose an integer of size approximately N, on average about how many distinct prime factors do we expect N to have (as $N \to \infty$)?
It might be useful to recall the Prime Number Theorem: The number of primes at most x is about $x / \log x$.

- (a) Around 6.
- (b) Around $\log \log \log N$.
- (c) Around $\log \log N$.
- (d) Around $\log N$.
- (e) Around $\log N \log \log N$.
- (f) Around $(\log N)^2$.
- (g) This is an open question.

Fermat Primes

Fermat Primes

If $F_n = 2^{2^n} + 1$ is prime, we say F_n is a Fermat prime. About how many Fermat primes are there less than x as $x \to \infty$?

Fermat Primes

Fermat Primes

If $F_n = 2^{2^n} + 1$ is prime, we say F_n is a Fermat prime. About how many Fermat primes are there less than x as $x \to \infty$?

- (a) 5
- (b) 10
- (c) Between 11 and 20.
- (d) Between 21 and 100.
- (e) $\log\log\log x$.
- (f) $\log\log x$.
- (g) $\log x$.
- (h) More than $\log x$.
- (i) This is an open problem.

$3x + 1$ Problem

$3x + 1$: Iterating to the fixed point

Define the $3x + 1$ map by $a_{n+1} = \frac{3a_n+1}{2^k}$ where $2^k || 3a_n + 1$. Choose a large integer N and randomly choose a starting seed a_0 around N. About how many iterations are needed until we reach 1 (equivalently, about how large is the smallest n such that $a_n = 1$)?

$3x+1$ Problem

$3x+1$: Iterating to the fixed point

Define the $3x+1$ map by $a_{n+1} = \frac{3a_n+1}{2^k}$ where $2^k || 3a_n+1$. Choose a large integer N and randomly choose a starting seed a_0 around N. About how many iterations are needed until we reach 1 (equivalently, about how large is the smallest n such that $a_n = 1$)?

There is a constant C so that the answer is

- (a) Around 6.
- (b) Around $C \log \log \log N$.
- (c) Around $C \log \log N$.
- (d) Around $C \log N$.
- (e) Around $C \log N \log \log N$.
- (f) Around $C(\log N)^2$.
- (g) This is an open question.

Expectation

Definition

Moments

Let X be a random variable. We define

- k^{th} moment: $m_k := \mathbb{E}[X^k]$ (if converges absolutely).

Assume X has a finite mean, which we denote by μ (so $\mu = \mathbb{E}[X]$). We define

- k^{th} centered moment: $\sigma_k := \mathbb{E}[(X - \mu)^k]$ (if converges absolutely).

- Be alert: Some books write μ'_k for m_k and μ_k for σ_k.
- Call σ_2 the variance, write it as σ^2 or $\mathrm{Var}(X)$.
- Note $\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Key Results on Expected Values

- Linearity: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
- Independence: X, Y independent then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. If RHS holds say uncorrelated.
- Variance: $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$ if uncorrelated. In general:

$$
\begin{aligned}
\mathrm{CoVar}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{CoVar}(X_i, X_j).
\end{aligned}
$$

Solutions
(Don’t read before the talk!)

Cereal toy problem

The answer is is (c), about $N \log N$. Let X_k denote how long we must wait till we get the k^{th} new toy, given that we have k distinct toys. Then if X is the total waiting time, $X = 1 + X_2 + \cdots + X_n$, and $X_k \sim \text{Geom}\left(\frac{n-(k-1)}{n}\right)$. It is a standard result that a geometric random variable with parameter p has expected value $1/p$. Inputting this leads to $\mathbb{E}[X] = \sum_{k=1}^{n} \frac{n}{k} \sim n \log n$.

An excellent challenge is to figure out a formula for the median waiting time (i.e., how long we must wait until we have a 50% chance of having one of each toy).

See also the solution to Problem 2 from Section 3.3 (page 13) in `http://www.williams.edu/go/math/sjmiller/public_html/341/handouts/hwcomments.pdf`.

Number of prime divisors

The answer is (c), around $\log \log N$. This is explained in detail in the supplemental notes to the lecture, online at http://www.williams.edu/go/math/sjmiller/public_html/wellesley/ExpectationThursOct8_2009

Number of prime divisors (continued)

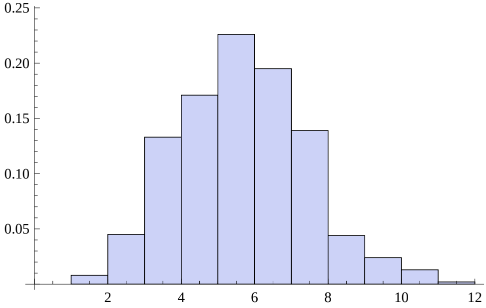

Figure: Distribution of the number of prime factors for n, 1000 consecutive values starting at $a_0 = 5487525252462375634352364513298043621345687989991218989811$. Note $\log \log a_0 \approx 4.88$.

Fermat Primes

We expect there to only be five Fermat primes, so (a). For a generic number N, the probability it is prime is about $1/\log N$. Thus the expected number of Fermat numbers that are prime should be

$$\sum_{n=0}^{\infty} \frac{1}{\log F_n} \approx \sum_{n=0}^{\infty} \frac{1}{2^n \log 2} \approx 2.88;$$

of course, we know the first 5 choices of n yield primes.... (Note: it isn't too surprising that we can have small discrepancies when the final answer is finite.)

$3x + 1$

The answer is (d), around $C \log N$ (it turns out C is about $1/\log(4/3)$). Let $x_n = \log a_n$. We have

$$\mathbb{E}[x_{n+1}] = \sum_{k=1}^{\infty} \frac{\log(3a_n + 1)}{2^k} \approx \sum_{k=1}^{\infty} \frac{x_n + \log 3}{2^k};$$

after some algebra we find the right hand side is $x_n + \log(3/4)$. Iterating we find $a_n \sim (3/4)^n a_0$, so the number of iterations expected before a_0 decays to 1 should be found by setting $(4/3)^n$ equal to a_0, so $n \sim \frac{\log a_n}{\log(4/3)}$.