A Statistical Look at Maps of the Discrete Logarithm

Joshua Holden and Nathan Lindle, Rose-Hulman Institute of Technology

Several algorithms in cryptography are based on the apparent difficulty of solving the discrete logarithm. It, like integer factorization, is attractive in cryptography because the inverse (modular exponentiation) is much easier to compute. The paper "Mapping the Discrete Logarithm" by Daniel Coulter and Joshua Holden takes a look at the functional graphs that can be generated using the function $x \mapsto g^x \pmod{p}$ where p is a prime number. It turns out the structure of the graph is largely determined by the interaction between g and $p-1$, and this interaction gives us an easy way to generate many binary functional graphs by choosing the correct values for g. In "Mapping the Discrete Logarithm" the authors extracted some statistics from graphs with p near 100,000. Their work showed evidence that binary functional graphs generated through modular exponentiation were very close to the theoretical values for random binary functional graphs.

This second look tells a slightly different story though. Using many of the same techniques as in the previous paper, we were able to generate values for the theoretical variance in several of the statistics which were measured. While the variance in the number of components and the number of cyclic nodes is similar to the theoretical variance for binary functional graphs, the variance in the average cycle length and the average tail length are far from what was expected. T-tests were also performed to determine if the theoretical and observed means were statistically similar. The results show that in some cases this is true, but in others the test shows that seemingly small differences could actually be quite significant. We have also applied a t-test on the variance to determine the significance of the deviation from what we expected.

Through some optimizations to the code used for the previous paper, as well as a conversion from C++ to C, we are now able to complete trials much more quickly. This has allowed us to test values of p near 200,000, and we have run tests on 33 primes between 100,00 and 200,00. The following is a tabulation of some of the statistics we have gathered, along with the variation, theoretical variation, and the P-value obtained after running a two-tailed t-test on the observed statistic comparing it to the theoretical value. The results here (unusual average cycle variation, average tail variation and maximum tail) are consistent with the rest of our results.

	100043			100057		
	Predicted	Observed	P-value	Predicted	Observed	P-value
Components	6.392	6.389	0.842	6.392	6.364	0.134
Variance	5.158	5.117	0.230	5.158	5.098	0.368
Cyclic nodes	395.417	395.303	0.920	395.445	395.858	0.842
Variance	42543.192	42227.348	0.272	42549.173	42781.153	0.690
Avg cycle	198.208	198.319	0.920	198.222	198.215	1
Variance	27210.527	**20392.727**	0	27214.349	**20680.648**	0
Avg. tail	197.212	197.178	1	197.226	196.768	0.548
Variance	27210.956	**7362.882**	0	27214.778	**7335.733**	0
Max cycle	247.495	247.261	0.764	247.512	247.302	0.920
Variance	NA	23806.218	NA	NA	23985.249	NA
Max tail	547.935	541.827	0	547.974	541.701	0
Variance	NA	26848.354	NA	NA	23985.249	NA

Table 1: Observed and theoretical statistics for p=100043 and p=100057

	106261			200087		
	Predicted	Observed	P-value	Predicted	Observed	P-value
Components	6.422	6.370	0.022	6.738	6.745	0.368
Variance	5.188	5.176	0.920	5.505	5.517	0.690
Cyclic nodes	407.551	408.433	0.690	559.620	562.252	0.006
Variance	45199.846	44488.375	0.272	85318.241	85825.849	0.230
Avg. Cycle	204.275	206.612	0.110	280.310	281.659	0.046
Variance	28907.991	**22003.465**	0	54537.238	**41358.233**	0
Avg. Tail	203.279	201.644	0.058	279.313	278.974	0.424
Variance	28908.420	**7578.376**	0	54537.668	**14731.689**	0
Max cycle	255.070	256.986	0.230	350.012	351.356	0.058
Variance	NA	25629.420	NA	NA	48068.838	NA
Max tail	564.756	554.905	0	775.570	769.207	0
Variance	NA	27602.038	NA	NA	54039.057	NA

Table 2: Observed and theoretical statistics for p=200087 and p=106261