Analysis of the quadratic seive

Say that we have an integer $n = pq$ (p and q primes $\neq 2$ and distinct). What we would like to do is choose k different integers in the range of $[1, (n-1)/2]$ and determine what is the probability that either (a) one of these integers has a a factor in common with n or (b) two integers have the same square modulo n.

To calculate this we require the following fact:

Proposition: If a is a quadratic residue $(mod\ pq)$ (that is, if there exists a integer y such that $y^2 \equiv a\ (mod\ pq)$) such that $gcd(a, pq) = 1$, then there are exactly 4 integer solutions (between 0 and $pq - 1$) to the equation

$$x^2 \equiv a\ (mod\ pq).$$

Note that because if x is a solution to $x^2 \equiv a\ (mod\ pq)$, then $-x$ will also be a solution to this equation, hence there are exactly two solutions in the range of $[0, (pq-1)/2]$.

Proof:

We require two observations in order to justify this proposition. The first says that
Observation 1: If $a \neq 0$ is a quadratic residue $(mod\ p)$, then there are exactly two solutions to the equation $x^2 \equiv a\ (mod\ p)$.

We of course know that modulo p there are at most r solutions to a polynomial equation of degree r, what we are observing here is that there are exactly 2 for a quadratic equation, d and $-d$ and these are not equal for p and odd prime.

Observation 2: Let $0 \leq a_1 < p$ and $0 \leq a_2 < q$, then there exists a unique $a \in [0, pq-1]$ such that

$$a \equiv a_1\ (mod\ p)$$

$$a \equiv a_2\ (mod\ q)$$

This is just a restatement of the Chinese remainder theorem in a way that emphasizes the uniqueness of the solution. There are $p \cdot q$ integers in the interval $[0, pq-1]$ and there are $p \cdot q$ pairs of integers (a_1, a_2) with $a_1 \in [0, p-1]$ and $a_2 \in [0, q-1]$.

Assume that $gcd(a, pq) = 1$ is a quadratic residue $(mod\ pq)$, then if x is an integer such that $x^2 = a\ (mod\ pq)$, then $x^2 \equiv a\ (mod\ p)$ and $x^2 \equiv a\ (mod\ q)$ each have exactly two solutions (since a is not 0 modulo p or q) $x \equiv \pm a_1\ (mod\ p)$ and $x \equiv \pm a_2\ (mod\ q)$. These four values for x modulo p and modulo q uniquely determine x modulo pq by our second observation. ⊓⊔

The birthday paradox

We remind the reader of the birthday paradox because it is relevant to what we will do next.

Say that there are 36 people in a room (I choose 36 because it is roughly one tenth of 365).
What is the probability that at least two of them have the same birthday?

Is it approximately ?
a) 10%
b) 20%
c) 50%
d) 85%

It seems quite surprising that the answer is very close to 85%. Our intuition about probability seems to tell us that the probability should be close to 1/10 or maybe 2/10. It is even more surprising to learn that if there are 50 people in the room then the probability that at least two have the same birthday is more than 97%. The way that we determine this is by calculating instead the probability that everyone in the room has a different birthday and then noting that

$$P(\text{ that at least 2 of the 36 have the same birthday }) = 1 - P(\text{ all 36 people have different birthdays })$$

Say that there are k people in the room, to calculate the probability that all k of them have different birthdays we take each one of these people one by one and cross their birthday off the calendar and say that the probability of the event that the r^{th} person has a birthday different from all the predecessors is

$$\frac{\text{number of days which are not the other } r-1 \text{ birthdays}}{\text{number of days in the year}} = \frac{365 - r + 1}{365}$$

This implies that

$$P(\text{all } k \text{ people have different birthdays}) = \frac{365}{365}\frac{364}{365}\cdots\frac{365-k+1}{365}$$

Factoring integers

Now we would like tackle the problem of factoring a large integer $n = pq$ by a parallel computation of choosing k integers in the range of $[0, (n-1)/2]$ and looking for either two integers with the same square modulo n (so that $a^2 - b^2 \equiv (a+b)(a-b) \equiv 0 \ (mod\ pq)$ and $gcd(a+b, n) > 1$) or more simply that $gcd(a, n) > 1$. For a given number n, how big do we need to choose k before we can expect to find a factor? We note that

$$P(\text{ that at least two of the } k \text{ integers have the same square or a factor in common with } n \text{ }) =$$
$$1 - P(\text{ all } k \text{ integers have different squares and are relatively prime to } n \text{ })$$

We proceed as we did in the birthday paradox and take each of these integers in the range $[0, (n-1)/2]$ and cross off both that integer and the one that has the same square. Then to find the probability of the event that the r^{th} integer is relatively prime to n and has a different square than all of its predecessors we want it to be one of the $\phi(n)/2 - 2r + 2$ integers from the $(n-1)/2 - r + 1$ integers that are remaining. Therefore the probability that the r^{th} integer is relatively prime to n and has a different square than all of the predecessors is

$$\frac{\text{number of relatively prime integers which don't match the other } r-1 \text{ squares}}{\text{total number of integers in } [0,(n-1)/2] \text{ which differ from the } r-1 \text{ previous}} = \frac{\phi(n)/2 - 2r + 2}{(n-1)/2 - r + 1}$$

$$P(\text{all } k \text{ integers have different squares and are relatively prime to } n) =$$
$$\frac{\phi(n)/2}{(n-1)/2}\frac{\phi(n)/2-2}{(n-1)/2-1}\frac{\phi(n)/2-4}{(n-1)/2-2}\cdots\frac{\phi(n)/2-2k+2}{(n-1)/2-k+1}$$

If we are trying to factor n, then we don't know the the value of $\phi(n)$ but we can determine an approximate value. For instance when RSA-129 was factored it was known that the integer was the product of a 64 and a 65 digit prime. To make an estimate of these values one could take $\phi(n) \approx n - 10^{65} - n/10^{65} + 1$ (which is still rather difficult to calculate).