Problem 5: Abnormal Gauss [HackerRank]

By Sacheth Sathyanarayanan

This problem was probably one of the harder problems on the contest, with only one correct submission. It was derived from a paper on efficient sampling from the normal distribution. What's particularly beautiful about this procedure is that it samples from the standard normal distribution while only using integer arithmetic! (In the implementation we gave you, this was not completely the case, but it can be expanded to only include integer operations.) This is quite remarkable, since we saw that the probability density function of the standard normal has irrational (in fact, transcendental) numbers like π and e in it.

The main idea here is to split the computation of a normally distributed variable y into the computation of it's integer part k and the fractional part x.

We first sample over all integers, selecting k with probability $e^{-k/2}(1-1/\sqrt{e})$. We can do this by randomly generating Bernoulli random variables with success probability $1/\sqrt{e}$. The required probability equals the probability of having exactly k consecutive successes. How do we generate these Bernoulli random variables? The function `bern` helps us out here — more on that later.

Once we sample the k, we accept it with probability $e^{-m/2} = e^{-k(k-1)/2}$. This tells us that $m = \mathtt{g(k)} = \mathtt{k(k-1)}$. We now sample x in $\mathcal{U}[0,1]$ and accept it with probability $e^{-x(2k+x)/2}$ so that the relative probability of choosing a given pair of k and x is $e^{-(k+x)^2/2}$, which is exactly what we wanted. Now, with probability $1/2$, we return $y = k + x$ and with probability $1/2$, we return $-y = -k - x$ and we are done.

Now in order to be able to sample k, we need a process to generate a bernoulli random variable with success probability $1/\sqrt{e}$. This is exactly what `bern` does.

We keep generating uniform random variables (after first generating $1/2$) while it is greater than the last. We claim that the probability that we generate an even number of such random variables equals $1/\sqrt{e}$. Why? Well, we first claim that

Claim. If we carry out the process in `bern`, then the probability that do at least n iterations of the loop before we terminate is exactly $(2^n n!)^{-1}$.

Proof of Claim: We must have $1/2 < U_1 < U_2 < \cdots < U_n$ if and only if we carry out at least n iterations. This holds if and only if all the U_i are greater than $1/2$ and $U_1 < U_2 < \cdots < U_n$.

COS CON

But these two occurrences are actually independent! (Think about it: if I asked you to guess the ordering of the U_i and told you that they were all greater than $1/2$, would that actually give you any useful information?) Thus, the probability of our desired occurrence happening is just the product of the probabilities of each of the U_i being greater than $1/2$ and the probability of $U_1 < U_2 < \cdots < U_n$ holding. The former happens with probability 2^{-n}, and the latter happens with probability $1/n!$ by a symmetry argument: every ordering (of which there are $n!$) of the U_i should be equally likely. (Since we are drawing our samples from a continuous distribution, we don't have to worry about the case where any two of the U_i are equal — this happens with probability zero.) $\square$

As such, the probability that we execute *exactly* n iterations is

$$\frac{1}{2^n n!} - \frac{1}{2^{n+1}(n+1)!}$$

And if we sum this over all even n, we get

$$\sum_{n \text{ even}} \left(\frac{1}{2^n n!} - \frac{1}{2^{n+1}(n+1)!} \right) = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2^2 \cdot 2!} - \frac{1}{2^3 \cdot 3!}\right) + \left(\frac{1}{2^4 \cdot 4!} - \frac{1}{2^5 \cdot 5!}\right) + \cdots$$

which we can notice is *exactly* the Taylor series for e^{-x} with $x = 1/2$ plugged in!

Thus, we should have that `f(v,n)` = `(n % 2 == 0)`. Once we know what `f` and `g` are, returning the value of the required expression is straightforward.

Plaudits:

- Congratulations to Calvin Deng (grad) for being the only participant to correctly solve this problem!