

# Optimal stopping rule for quantum computer

Denys Gagnon

June 3, 2022

## 1 Summary

We present an optimal method for stopping iterations on a quantum computer. We first consider the simple case of estimating  $\alpha^2$ . Then we look at the estimation of a differentiable function  $f(\alpha^2)$  and apply it to the case  $\sqrt{\alpha^2} = \alpha$ . We then look at estimation for several qubits at the same time. Lastly we consider the problem in the presence of random noise.

## 2 Estimation for a single qubit

When a qubit in state  $\alpha|0\rangle + \beta|1\rangle$  is measured, it either returns 0 or 1 with probability  $p = \alpha^2$  and  $1 - p = \beta^2 = 1 - \alpha^2$ . The objective of a quantum computer program is often to get the value of  $\alpha$ . So the program is run a large number of time and the relative frequency of zeros is computed. By the law of large numbers that proportion will converge to  $\alpha^2$ . The number of time the programs needs to run depends on the desired precision of the estimate. This precision is defined as a interval of confidence. That is, a maximum distance between the real value of  $\alpha$  and the value of the estimate with a minimum probability. This can be rephrase as: *the estimate does not differ from the real value by more than  $\Delta$  with a probability greater than  $1 - \gamma$* . The parameters  $\Delta$  and  $\gamma$  are fixed by the program user and the program is run until the precision is met.

Unfortunately, the number of required iterations is dependant on the unknown value so the user is left to do trial until the result is satisfactory.

In the following paper we presents a method to sequentially assess is a sufficient number of iterations have been done to meet the precision requirements. The users simply enters the precision parameters  $\Delta$  and  $\gamma$  and let the program iterate until it is reached.

In the first section we present the methods for estimating  $\alpha^2$  for one qubit at a time. We then address the case for  $f(\alpha^2)$ . In the next section we address the problem of estimating several qubits at the same time. We then consider the problem of noise in the estimation process.

### 2.1 Sampling for $\alpha^2$

Suppose the qubit is measured  $N$  times and let  $\{X_1, X_2, \dots, X_N\}$  be the observed values. We have that  $P(X_n = 0) = |\alpha|^2 = p$ , hence the following natural estimator of  $|\alpha|^2$ :

Let  $S_N = X_1 + \dots + X_N$ ,  $F_N = N - S_N$  and

$$\hat{p} = \widehat{|\alpha|^2} = \frac{S_N}{N} \quad (1)$$

The variance of this estimator is given by:

$$V(\hat{p}) = \frac{p(1-p)}{N}$$

If  $M$  is large, then

$$Z = \sqrt{\frac{N}{p(1-p)}} (\hat{p} - p) \quad (2)$$

has an approximate normal distribution with mean 0 and variance 1.

If one is interested in getting a estimate of  $p$  with a fixed minimal precision, it will usually be in the form of a interval  $[\hat{p} - \Delta, \hat{p} + \Delta]$  havin a probability  $1 - \gamma$  of containing the true unknown  $p$ . Typically,  $\Delta$  is defined by the number of significant digits:  $\Delta = 2 \times 10^{-k}$ , where  $k$  is the number of digits. The confidence level is defined by a small value of  $\gamma$  such as 5%, 1% or even less. The smaller the value of  $\gamma$  the more likely the probability for the interval to include the real, but unknown, value of  $p$ .

Using the asymptotic normal distribution of  $Z$  in (2), let  $\Phi()^{-1}$  be the inverse of the normal cumulative distribution function and set

$$z = \Phi^{-1}(1 - \gamma/2)$$

We seek  $N$  such that

$$\begin{aligned} 1 - \gamma/2 &= P \left[ -z < \sqrt{\frac{N}{p(1-p)}} (\hat{p} - p) < z \right] \\ &= P \left[ -z \sqrt{\frac{p(1-p)}{N}} < (\hat{p} - p) < z \sqrt{\frac{p(1-p)}{N}} \right] \end{aligned}$$

The length of the interval is therefore given by  $2z \sqrt{p(1-p)/N}$  and must be less or equal to  $2\Delta$ , hence:

$$N = p(1-p) \left[ \frac{\Phi^{-1}(1 - \gamma/2)}{\Delta} \right]^2 \quad (3)$$

where the constant  $\Phi^{-1}(1 - \gamma/2)/\Delta$  is large given that  $\Phi^{-1}(1 - \gamma/2)/$  is larger than 1 and that  $\Delta$  is usually a small number accounting for the number of correct digits.

The problem with equation (8) is that it relies on the knowledge of  $p$  which is precisely what we are trying to find. One option is to take the worst case, which happens when  $p = 1/2$ . In this case (8) becomes:

$$N = \left[ \frac{\Phi^{-1}(1 - \gamma/2)}{2\Delta} \right]^2$$

As an example, with  $\Delta = 0.0001$  and  $\gamma = 0.05$ , we have  $N = 95,062,500$ .

An alternate solution is to go sequentially and update the target number of trial as we get a better estimate of  $N$ .

## 2.2 Sequential Estimation

If we proceed sequentially, the sample size is increased progressively and, at each step, a decision is taken as to continue or stop.

It would be tempting to simply reuse equation 8 with  $\hat{p}$  estimated after a few steps to compute  $N$  with:

$$N = \hat{p}(1 - \hat{p}) \left[ \frac{\Phi^{-1}(1 - \gamma/2)}{\Delta} \right]^2 = \frac{S_N F_N}{N^2} \left[ \frac{\Phi^{-1}(1 - \gamma/2)}{\Delta} \right]^2$$

Setting

$$H(\gamma, \Delta) = \left( \frac{\Phi^{-1}(1 - \gamma/2)}{\Delta} \right)^{2/3} \quad (4)$$

we have:

$$N = (S_N F_N)^{1/3} H(\gamma, \Delta) \quad (5)$$

On average,  $(S_N F_N)^{1/3}$  increases at rate  $N^{2/3}$ . Hence, at one point  $N$  will become greater and sampling should stop.

There is one problem however with equation 5. At the beginning of the process it is very likely to get several zeros in a row without ones (or the reverse). In that case,  $\sqrt{S_N F_N}$  is null and the process should stop immediately. This situation is more likely to happen when  $p$  is close to either 1 (or 0 for the reverse situation).

Setting a minimal number of trial before we start comparing  $N$  with  $S_N F_N$  would be an option. In order to determine such a value we must consider the case where  $p$  is very small and remember that is are satisfied with a precision of  $\Delta$  for our estimate. So any value of  $p$  estimated to be 0 while it is in fact smaller than  $\Delta$  is acceptable.

Assume now that  $p = \Delta$ , so that the probability that  $X_n = 1$  is very high. Then the probability of getting a sequence of  $N$  ones in a row is given by  $(1 - \Delta)^N$ . We don't want to make the error of claiming the  $p = 0$  with a higher probability than  $1 - \gamma$ . Therefore,  $N$  must be large enough so that  $\Delta^N < 1 - \gamma$ . Hence,  $N > \log(1 - \gamma) / \log(1 - \Delta)$ . For  $p > 1 - \Delta$  the result is the same.

The final stopping rule is therefore:

$$\text{Stop samplig when } N \geq \max \left[ \frac{\log(1 - \gamma)}{\log(1 - \Delta)}, (S_N F_N)^{1/3} H(\gamma, \Delta) \right] \quad (6)$$

With  $\Delta = 0.0001$  and  $\gamma = 0.05$ , we have  $H(0.05, 0.0001) = 727$  so :

$$\text{Stop samplig when } N \geq \max [9, 500, 727 (S_N F_N)^{1/3}]$$

Since  $S_N F_N \approx N^2 p(1 - p)$ , for  $p = 1/2$  at trial 9,500 we have  $727 (S_N F_N)^{1/3} \approx 727 (9,500^2/4)^{1/3} = 205,430$ . In the worst case  $S_N$  and  $F_N$  are very close ( $\approx N/2$ ) and the process will go on until  $N = 727 (N^2/4)^{1/3}$  which happens when  $N = 727^3/4 = 96,060,145$  which is very close the worst case with a fixed sample size.

For  $p = 1/4$ ,  $S_N F_N \approx 3N^2/16$  and the process will likely stop for  $N$  in the neighbourhood of  $(3/16)727^3 = 72,045,109$ , less than 19% of the worst case.

### 2.3 Estimation of $f(|\alpha|^2)$

Often, the parameter that is to be measured is not  $\alpha^2$  but  $\alpha = \text{sqrt}(\alpha^2)$  or a differentiable function  $f$  of it. If  $f$  is continually differentiable, then  $\sqrt{N}(f(\hat{p}) - f(p))$  has an asymptotic normal distribution with mean 0 and variance  $f'(p)^2 p(1 - p)$ . So, using the same argument as in the previous section:

$$\begin{aligned} 1 - \gamma/2 &= P \left[ -z < \sqrt{\frac{N}{f'(p)^2 p(1 - p)}} (f(\hat{p}) - f(p)) < z \right] \\ &= P \left[ -z \sqrt{\frac{f'(p)^2 p(1 - p)}{N}} < (\hat{p} - p) < z \sqrt{\frac{f'(p)^2 p(1 - p)}{N}} \right] \end{aligned} \quad (7)$$

The length of the interval is therefore given by  $2z \sqrt{f'(p)^2 p(1 - p)/N}$  and must be less or equal to  $2\Delta$ , hence:

$$N = f'(p)^2 p(1 - p) \left[ \frac{\Phi^{-1}(1 - \gamma/2)}{\Delta} \right]^2 \quad (8)$$

We can now use the same sequential method

$$N = f'(\hat{p})^2 \hat{p}(1 - \hat{p}) \left[ \frac{\Phi^{-1}(1 - \gamma/2)}{\Delta} \right]^2 = f' \left( \frac{S_N}{N} \right)^2 \frac{S_N F_N}{N^2} \left[ \frac{\Phi^{-1}(1 - \gamma/2)}{\Delta} \right]^2$$

$$N = f' \left( \frac{S_N}{N} \right)^{2/3} (S_N F_N)^{1/3} H(\gamma, \Delta) \quad (9)$$

And samplig should continue until  $N$  is not greater than the right hand side of 9.

As for the simpler case of estimating  $\alpha^2$ , we need to deal with the cases of a long sequence of zeros, or ones which could happen when  $p = 0$  or  $p = 1$ , but also when  $f'(p)^2 p(1-p)$  is small. But the way we solved the first situation also resolved the problem of a very small derivate. Since we'll be doing a minimal number of trials equal to  $\log(1-\gamma)/\log(1-\Delta)$ , then we know that, with probability no less than  $1-\gamma$ , that  $|\hat{p} - p| < \Delta$ . And if  $f'(p)^2 \leq 1$ , then  $|f(\hat{p}) - f(p)| < \Delta$  with the same probability. We therefore have the necessary precision for small values of  $f'(p)^2$ .

The stopping rule becomes:

$$\text{Stop samplig when } N \geq \max \left[ \frac{\log(1-\gamma)}{\log(1-\Delta)}, f' \left( \frac{S_N}{N} \right)^{2/3} (S_N F_N)^{1/3} H(\gamma, \Delta) \right]$$

For example, with  $f(p) = \sqrt{p}$ , we have:

$$f' \left( \frac{S_N}{N} \right)^{2/3} (S_N F_N)^{1/3} = N^{1/3} \frac{1}{(4S_N)^{1/3}} (S_N F_N)^{1/3} = \left( \frac{N}{4} \right)^{1/3} F_N^{1/3}$$

One last improvement could be to get rid of the derivative and estimate it instead of requiring the user to provide a function to do it. Let

$$D_N = (f(S_N/N + \Delta) - f(S_N/N - \Delta))/2\Delta$$

Then we have:

$$\text{Stop samplig when } N \geq \max \left[ \frac{\log(1-\gamma)}{\log(1-\Delta)}, D_N^{2/3} (S_N F_N)^{1/3} H(\gamma, \Delta) \right]$$

### 3 Simultaneous estimation of several qubits

If dealing with entangled qubits the measure is made on a multinomial distribution. For instance, the two qubits  $|\psi_1\psi_2\rangle$  are in state  $\alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$ , where  $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$ .

Suppose  $K$  qubit are to be measured. Then, there are  $L = 2^K$  possible measurement outcomes. The parameter  $p$  is not a scalar anymore but a  $2^K$ -dimensional vector  $p_1, \dots, p_L$ , representing the probability of each of the possible outcomes.

Let  $S_{lN}$  the proportion of time the the outcome  $l$  appeared and  $\hat{p}_l = S_{lN}/N$ . Of course  $\sum_l S_{lN} = N$  and  $\sum_l \hat{p}_l = 1$ . The variance of  $\hat{p}_l$  is equal to  $p_l(1-p_l)/N$  and the correlation between  $\hat{p}_l$  and  $\hat{p}_m$  is given by  $-p_l p_m/N$ .

The probability that all outcome estimate will be within less that  $\Delta$  from their true values is difficult to compute exactly given the correlation. The simplest alternative is to use the Bonferonni method. If a global probability of  $1-\gamma$  is desired then the individual probability has to be set as  $1-\gamma/L$  and use  $H(\gamma/L, \Delta)$  (4) with the folloing rule:

$$\text{Stop samplig when } N \geq \max \left[ \frac{\log(1-\gamma/L)}{\log(1-\Delta)}, \max_l (S_{lN} F_{lN})^{1/3} H(\gamma/L, \Delta) \right] \quad (10)$$

When estimating  $f(p_1, \dots, p_L)$ , where  $f : \mathfrak{R}^L \rightarrow \mathfrak{R}$ . Then

$$f(\hat{p}_1, \dots, \hat{p}_L) - f(p_1, \dots, p_L) = \frac{\partial f}{\partial p_1}(p)(\hat{p}_1 - p_1) + \dots + \frac{\partial f}{\partial p_L}(p)(\hat{p}_L - p_L)$$

Hence

$$V(f(\hat{p}_1, \dots, \hat{p}_L)) = \left[ \frac{\partial f}{\partial p_1}(p) \right]^2 V(\hat{p}_1) + \dots + \left[ \frac{\partial f}{\partial p_L}(p) \right]^2 V(\hat{p}_L)$$

## 4 Sampling with external white noise

So far we have assumed that, when measuring the qubit, the random variable  $X_n$  had a binomial distribution with parameter  $p = |\alpha|^2$ . This will happen if there is no external noise disturbing the qubit. If this is not the case we can assume that the impact of the external noise is not correlated with the value of the qubit when it switches the lecture. In order to model this we introduce a new random variable  $Y_n$  that may change the value of the reading with probability  $1 - q$ . If  $Y_n = 0$ , then the reading is exact and equals  $X_n$ . But if  $Y_n = 1$  then output of the qubit is the opposite of the correct value. It will be 1 if  $X_n = 0$  and 0 if  $X_n = 1$ .

$$\begin{array}{cc} & \begin{array}{cc} X_n = 0 & X_n = 1 \end{array} \\ \begin{array}{c} Y_n = 0 \\ Y_n = 1 \end{array} & \left[ \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right] \end{array} \quad (11)$$

Figure 1: Measured output  $Z_n$  depending on  $X_n$  and  $Y_n$

The impact of  $Y_n$  would be minimal on  $X_n$  if  $p = 1/2$  since the change from 0 to 1 would be, on average, offset by the changes from 1 to 0. However, if  $p$  is closer to 0 or to 1, then there is a higher probability the one type of change will happen more frequently than the other.

We have  $P(Z_n = 0) = P(X_n = 0 \cap Y_n = 0) + P(X_n = 1 \cap Y_n = 1) = pq + (1-p)(1-q) = pq + 1 - q - p + pq = 1 + 2pq - p - q = p(2q - 1) + (1 - q)$ . As we can see, there is a bias introduced in the measurement. So a simple average of the  $Z_n$  will lead to an estimation of  $p(2q - 1) + (1 - q)$  instead of the desired  $p$ . If  $q$  is known then

$$\hat{p} = \frac{N^{-1} \sum_{n=1}^N Z_n - (1 - q)}{(2q - 1)} \quad (12)$$

is an unbiased estimator for  $p$  and preferable to (1). If the noise is low,  $q$  will be close to 1 and  $1 - q$  near 0 and negligible, while  $2q - 1$  is almost equal to 1 and negligible as well.

The variance of  $Z_n$  is then equal to  $(p + q - 2pq)(1 - p - q + 2pq)$  and the variance of the estimator in (12) is:

$$V(\hat{p}) = \frac{(p + q - 2pq)(1 - p - q + 2pq)}{N(2q - 1)^2} \quad (13)$$

And the sample size can be treated similarly to the previous case as in (??). One can see that as  $q$  gets near 0.5 the variance will increase and require sample size much larger.

The value of  $q$  has to be known for (12) to be useful. This can be based on past experiments made on the qubit and could be considered a calibration step.

## 5 White noise increasing with time

In the previous section we have assumed that the white noise  $Y_n$  was a random variable with a constant probability  $q$  of altering the outcome of the measurement. In reality, qubits will progressively become noisier as the time goes by. The probability  $q$  is in fact a function of time  $q(t)$  which is 0 when  $t=0$  and progressively increases to 0.5.

A qubit is first initialised and then several calculations are done on it using logical gates. As these gates take a certain time to execute this must be added to the time of the qubit. If the time for each gate is a constant  $T_0$  then the time to perform the computation on the qubit is equal to the number of gates multiplied by the time used to execute the gate:  $q(t) = q(T_0 \times \text{Number of gates in the computation}) = q(T_0 K)$ . The function  $q()$

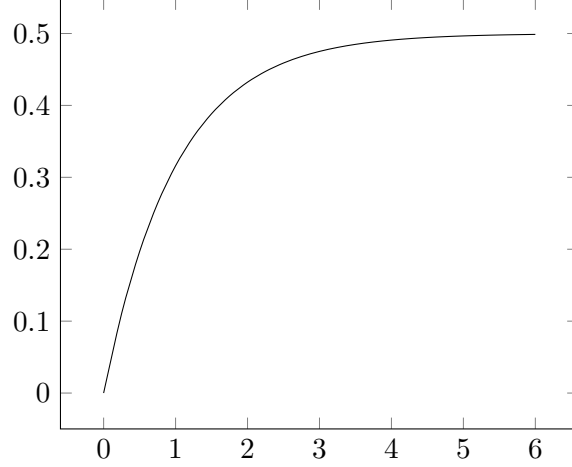


Figure 2:  $q(t)$  as a function of time

could be approximated by an exponential such as:

$$\begin{aligned}
q(t) &= 0.5(1 - e^{-C_0 t}) \\
&= 0.5(1 - e^{-C_0 T_0 K}) \\
&= 0.5(1 - e^{-B_0 K})
\end{aligned} \tag{14}$$

where  $B_0$  is a constant for a given qubit.

Inseting the last line of (14) into equations (12) and (13) we get:

$$\hat{p} = \frac{N^{-1} \sum_{n=1}^N Z_n - (1 - 0.5(1 - e^{-B_0 K}))}{(20.5(1 - e^{-B_0 K}) - 1)} \tag{15}$$

and

$$V(\hat{p}) = \frac{(p + 0.5(1 - e^{-B_0 K}) - 2p0.5(1 - e^{-B_0 K}))(1 - p - 0.5(1 - e^{-B_0 K}) + 2p0.5(1 - e^{-B_0 K}))}{N(20.5(1 - e^{-B_0 K}) - 1)^2} \tag{16}$$

$$V(\hat{p}) = \frac{(p + 0.5(1 - e^{-B_0 K}) - p(1 - e^{-B_0 K}))(1 - p - 0.5(1 - e^{-B_0 K}) + p(1 - e^{-B_0 K}))}{N((1 - e^{-B_0 K}) - 1)^2} \tag{17}$$

$$V(\hat{p}) = \frac{0.25 - (p - 0.5)^2 e^{-2B_0 K}}{N e^{-2B_0 K}} \tag{18}$$

## 6 Computer implementation

The formula mentionned above are used by the classical computer controlling the quantum computer. The circuit definition and measurement is send to the quantum computer by the classical one until the stopping rule is met. At that point the quantum computer is no more required.

From an end-user point of view the methods are implemented using a class that takes as an input

1. A methods submitting the circuit definition and measuremnts once and returning the result in the form of an array of zeros and ones (the bits).

2. The number of bits returned must be specified as a second argument.
3. A third argument is a function to be applied to the bits by the classical computer and returning a double precision number. If this argument is NULL then no function is applied.
4. The fourth argument is  $\Delta$  and
5. The fifth one is the value of  $\gamma$ .

As a helped the user may want to know what is the expected number of runs the process may take in the worst case and some intermediary cases. A standalone program is provided where the user can provide the information and a report is produced giving some insight as to what may happen, including execution time.