Presentation for the Quantum Seminar

José Manuel Rodríguez Caballero

University of Tartu

Spring 2020

Subject

My presentation is about the paper[1]:

> *Bouman, Niek J., and Serge Fehr. ”Sampling in a quantum population, and applications.” Annual Cryptology Conference. Springer, Berlin, Heidelberg, 2010.*

My presentation may differ in some points with respect of the presentation given by the authors.

[1] I will not repeat the notation from the paper in this presentation. If someone needs a clarification, either read the paper or ask me during the presentation.

Contributions of the paper

Contribution 1. The authors introduce a framework in for *sampling quantum population*.

Contribution 2. This framework is used in a new proof of the security of the *quantum key distribution protocol BB84* (entanglement-based version).

Contribution 3. This framework is used in a new proof of the security of the *quantum oblivious-transfer from bit-commitment*.

Presentation of Contribution 1.

The authors introduce a framework in for *sampling quantum population*.

Overview of Contribution 1

The authors define a *sampling and estimation strategy* (*sampling strategy* for short). This strategy can be used in order to study both classical and quantum populations. The “quantum error” is bounded by the square root of the “classical error”, thus reducing the quantum information-theoretical problem to a well-studied probabilistic problem.

Sampling strategy

Let n be a positive integer, $\mathcal{A}$ be a finite alphabet and $\mathcal{S}$ be a finite set of seeds. Assume that $|\mathcal{A}| \geq 2$ and $|\mathcal{S}| \geq 1$. A *sampling strategy* is a triplet (P_T, P_S, f), where P_T is a probability distribution over $\mathcal{T} := 2^{[n]}$, P_S is a probability distribution over $\mathcal{S}$ and f is a real-valued function defined on the finite set[2]

$$\mathrm{Dom}_f := \{(t, q, s) \in \mathcal{T} \times \mathcal{A}^* \times \mathcal{S} : \quad |t| = |q|\} .$$

Remark. The number of degrees of freedom of a sampling strategy is finite and it is asymptotically equivalent to $|\mathcal{S}|^2 \, (2|\mathcal{A}| + 2)^n$ as $n \to \infty$.

[2] The notation $\mathcal{A}^*$ is for the free monoid over $\mathcal{A}$.

Classical populations

What is a classical population?

In the context of the paper, a *classical population* is a finite word q over the alphabet $\mathcal{A}$. The alphabet $\mathcal{A}$ contains a distinguished element, that we will denote by 0. We are interested in estimating how many letters in q are different from 0, i.e., the Hamming weight of q. In order to simplify the presentation, we will use the relative hamming weight of q, denoted $\omega(q)$, which is the Hamming weight of q divided by its length.

Example of classical population

The classical population could be the population of the People's Republic of Wakanda[3] (each person corresponds to a position in the word q) and the alphabet could be $\mathcal{A} = \{0, 1\}$, where 0 and 1 stand for uninfected and infected of COVID-19, respectively. In order to estimate how many people are infested with COVID-19, we will study a random sample of the population consisting of k people, which will be small in comparison to the size of the population, i.e. k is "small" with respect to n. Finally, we will extrapolate the data from the sample to the whole population.

[3] I am considering an idealized situation, in order to avoid all the complications of the a real-life scenario, e.g., fake data and heterogeneous properties of the population. In Wakanda everything will be as homogeneous as possible.

How to estimate?

Given a sampling strategy (P_T, P_S, f) and a classical population $q = q_1 q_2 q_3 ... q_n \in \mathcal{A}^n$. We take a random pair $(t, s) \in \mathcal{T} \times \mathcal{S}$ (using the joint distribution of P_T and P_S). Let

$$\tau_1 < \tau_2 < ... < \tau_k$$

be the elements of t written in increasing order. Let

$$\overline{\tau}_1 < \overline{\tau}_2 < .... < \overline{\tau}_{n-k}$$

be the elements of $[n] \backslash t$ written in increasing order. We define

$$q_t = q_{\tau_1} q_{\tau_2} ... q_{\tau_k}, \quad q_{\overline{t}} = q_{\overline{\tau}_1} q_{\overline{\tau}_2} ... q_{\overline{\tau}_{n-k}}.$$

By definition, we *estimate* that the relative Hamming weight of $q_{\overline{t}}$ is given by $f\left(t, q_t, s\right)$.

Example of estimation

Let's return to the example of the population of Wakanda, identified with the word q. We define $\mathcal{S} = \{\perp\}$, i.e., we do not use a random seed here. So, P_S is trivially the uniform distribution on $\mathcal{S}$. Now, let assume that P_T is the uniform distribution on subsets of size k of $[n]$, i.e.,

$$P_T(t) = \frac{1}{\binom{n}{k}}$$

if $|t| = k$ and $P_T(t) = 0$ otherwise. Finally, define

$$f(t, q, s) := \omega(q).$$

This framework is known as *random sampling without replacement*.

Example of estimation (continuation)

This sampling strategy is just a formalization of the heuristic approach assuming that the frequency of cases of COVID-19 in the whole population is the same as in our uniformly random sample. This is not necessarily the case in reality, e.g., people working in hospitals have a higher probability of getting infested than the remain of the population.

Set of classical δ-close states

Let δ be a positive real number. We define the *set of classical δ-close states* as

$$B^{\delta}_{t,s} := \left\{ q \in \mathcal{A}^n : \quad \left| \omega \left(q_{\bar{t}} \right) - f \left(t, q_t, s \right) \right| < \delta \right\},$$

Heuristic idea: for δ, t and s fixed, the larger the set $B^{\delta}_{t,s}$ is, the better the function f is for estimation (with respect to t and s).

Example of set of classical δ-close states

In Wakanda example, $B^{\delta}_{t,s}$ is the set of all possibilities (after choosing t and s) in which the frequency of infection of COVID-19 in the sample is δ-close to the frequency of infection in the remaining part of the population of Wakanda.

Random set of classical δ-close states

Let T and S be random variables associated to the probability distributions P_T and P_S respectively. Notice that the pair (T, S) is a random variable. Furthermore, the evaluation of $(t, s) \mapsto B^{\delta}_{t,s}$ at (T, S) determines a random variable, denoted $B^{\delta}_{T,S}$ and associated to the probability distribution

$$\Pr\left[B^{\delta}_{T,S} \in \Gamma\right] := \Pr\left[(T, S) \in \left\{(t, s) \in \mathcal{T} \times \mathcal{S} : \quad B^{\delta}_{t,s} \in \Gamma\right\}\right]$$

where $\Gamma \subset 2^{\mathcal{A}^n}$. The call $B^{\delta}_{T,S}$ the *random set of classical δ-close states*.

Heuristic idea: for δ fixed, the "larger" the random set $B^{\delta}_{T,S}$ is, the better the strategy is for estimation.

Example of random set of classical δ-close states

In Wakanda example, $B^{\delta}_{T,S}$ is the random set of all possibilities (before choosing t and s) in which the frequency of infection of COVID-19 in the sample is δ-close to the frequency of infection in the remaining part of the population of Wakanda.

Classical δ-error

The *classical δ-error* is defined as

$$\varepsilon_c^\delta := \max_{q\in\mathcal{A}^n} \Pr\left[B_{T,S}^\delta \in \left\{X \in 2^{\mathcal{A}^n} : \quad q \notin X\right\}\right].$$

Heuristic idea: ε_c^δ measures the probability that the sampling strategy fails in the worst-case scenario[4].

[4] Notice that the expression inside the bracket is equivalent to $q \notin B_{T,S}^\delta$.

Example of classical δ-error

In Wakanda example, ε_c^δ is the largest probability, with respect to all possible populations, that the frequency of infection of the sample is not δ-close to the frequency of infection of the remaining part of the population of Wakanda.

For $k \leq \frac{n}{2}$ (as it should be in practice), we have

$$\varepsilon_c^\delta \leq 2 \exp\left(-\frac{\delta^2 k}{2}\right).$$

Remark: As expected: a) as the size of the sample increases, the bound of the error ε_c^δ decrease; b) as the precision of the estimation increases, the bound of the error ε_c^δ increases.

Quantum populations

Presentation of Contribution 2.

This framework is used in a new proof of the security of the *quantum key distribution protocol BB84* (entanglement-based version).

Presentation of Contribution 3.

This framework is used in a new proof of the security of the *quantum oblivious-transfer from bit-commitment*.

End of my presentation