ALBANIAN JOURNAL
OF MATHEMATICS
Volume 2, Number 3, Pages 259–265
ISSN 1930-1235: (2008)

QUANTUM CHANNELS WITH CONTINUOUS INPUT ALPHABET

Georges Parfionov

Dept. of Mathematics,
SPb EF University,
Griboyedova 30–32, 191023, St.Petersburg, Russia

Romàn R. Zapatrin

Dept. Informatics,
The State Russian Museum,
Inénernaya, 4, 191186, St.Petersburg, Russia
e-mail: Roman.Zapatrin@gmail.com

Abstract. Any communication assumes a preliminary agreement between the parties involved. In our paper we address the question: what can we get when there is no agreement between the parties in the framework of classical communication quantum channels. We admit the concept of mixed coding and starting from it derive an idealized communication scheme based on continuous coding.

1. Classical communication through quantum channel

For the sake of self-consistency, we start from the conventional scheme of classical communication through quantum channel. Its basic ingredients are:

- **Coding.** It contains
  - A set of input states associated with the symbols of input alphabet
  - For each input state its a priori probability π_j is given
- **Transmission.** It is described by a superoperator: an affine mapping from the state space of the input of the channel ti that of the output.
- **Receiving a signal.** It is described by applying appropriate measurement on the set of output states, so that:
  - A measurement is a resolution of unit
  - When a signal is received, we judge which was the input state
  - For each input state j we calculate the probabilities $p_M(k \mid j)$ of taking the decision that the observed symbol was k (for every k)
  - The task in *conventional framework* is to find an optimal procedure to decide which was the input state

When the input coding and the output measurements are fixed, the probability to take the right decision then reads:

$$P_M = \sum_j \pi_j p_M(j \mid j) \tag{1}$$

This kinds of tasks are typical for communication theory and mathematical statistics. Finding the maximum of P_M is called identification of signals based on the criterion of maximal likelihood. Given an input coding, the task is, varying the output measurement M, to find such one that the probability P_M defined in (1) becomes maximal.

In standard framework we are given an input ensemble, that is, a collection of input states ψ_j with given probabilities π_j. For the input ensemble, its average density matrix ρ is calculated:

$$\rho = \sum_j \pi_j \psi_j \tag{2}$$

What is crucial in this scheme is that the efficiency of the channel (1) primarily depends on the input ensemble rather than on its average density matrix (see, e.g. optimal coding schemes by Schumacher and Westmoreland [1])
What is peculiar for our framework. In our framework we suppose that *only the average* density matrix (2) of the input ensemble is known, while the ensemble itself is not given for us. For quantum mechanical systems there are (infinitely) many ensembles having the same average density.

In other words, we only know the channel as a physical system. Any communication assumes a preliminary agreement between the parties dealing with input and output of the channel. In our paper we address the question: what can we get when there is *no agreement* between the parties.

That means, we are given the state space of the input but we *do not know* the a priori probabilities of input states and the result of measurements reduces to specifying the average output density matrix.

The problems of this kind have a long history lasting from Laplace to Boltzmann; their are solved on the basis of the principle of maximal entropy.

2. From Laplace principle to maximal entropy

It was Laplace who introduced the *principle of insufficient reason*: if there is no reason to prefer one outcome w.r.t. another one, all outcomes are treated equally probable (provided they are mutually exclusive and collectively exhaustive). Its direct consequence was the formula of classical probability [2]:

$$p = \frac{\text{Favorables}}{\text{Possibles}} \tag{3}$$

According to Laplace, if we are given an unknown distribution and we *need* to estimate it, we assume it to be uniform.

But what should we do if we have an additional information about the distribution? Can we still use the Laplace principle?

Let us illustrate it on a classical example. Suppose we play with die whose properties are not known. If we are asked what is the probability of a face to appear, we intuitively (but in fact according to Laplace) answer 1/6.

Let N identical dice are rolled and the mean value M of the number of points appeared is known. First suppose it turned out to be 3.5. This is an additional information about the dice, and how it affects our estimation? In this particular case we see that result is compatible with the initial hypothesis:

$$M = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Now take another kind of 'biased' dice such that the appropriate average value turns out to be, say, 4. In this case the hypothesis of the equality of all *faces* is no longer compatible with initial hypothesis and the Laplace principle is not applicable, at least in its direct form.

Among this we have n_1 times face 1,..., n_6 times face 6. The values $n_1, \dots, n_6$ satisfy the equations

$$n_1 + \cdots + n_6 = N$$

and

$$1 \cdot n_1 + \cdots + 6 \cdot n_6 = M \cdot N$$

When we try to solve this system with respect to $n_1, \dots n_6$, we get many solutions. Although, each particular solution occurs with its frequency:

$$W(n_1, \dots, n_6) = \frac{N!}{n_1! \cdots n_6!}$$

We are finding the solution $n_1, \dots, n_6$, which has greatest probability to occur, therefore we maximize the value of the frequency $W(n_1, \dots n_6)$. As it is known (see, e.g. [3]):

$$\log W \sim N \cdot \left(-\frac{n_1}{N} \log \frac{n_1}{N} - \cdots - \frac{n_6}{N} \log \frac{n_6}{N} \right) \tag{4}$$

And the maximum is attained at

$$n_k \sim N \cdot \frac{e^{-\beta k}}{Z} \tag{5}$$

where the normalizing factor Z is

$$Z = \sum_k e^{-\beta k} \tag{6}$$

Although the knowledge of the mean value is an additional information, the Laplace principle still works and this particular mean value gives no preference to any state, therefore the null hypothesis (the uniform distribution $p_j = \frac{1}{6}$) should not be rejected.

Biased die. Now let us consider what happens when the average is 4. In this case the Laplace principle should be developed: namely, the distribution should have maximal entropy $H = -\sum p_j \log p_j$. In our particular case this gives the following answer:

$$p_j = \frac{e^{-\beta j}}{Z} \tag{7}$$

where $Z = \sum_j e^{-\beta j}$ and β is calculated from the given average value (in our case, 4)

$$\sum_j j \cdot p_j = 4$$

This principle extends the Laplace principle to the notion of maximal entropy [4].

Why the idea to maximize the entropy H is a development of Laplace idea of symmetry and non-preference? For any given average value we consider all possible distributions which yield this average value. Then the take such distributions which are typical, that is, which mostly occur in all possible configurations. The preference is given to what occur with maximal number of combinations. The statistical weight

$$W = \frac{N!}{\prod_j n_j}$$

where N is the total number of trials and n_j is number of occurrence of j-th face.

The main message of this section is the following. We provide a completely classical example where we have *no knowledge* about the input state (distribution) but we *need* to tell something about it. A principle is suggested to choose a concrete distribution on the basis of a given small amount of knowledge.

In the case of quantum systems these distributions will be of a particular kind — continuous ensembles.

3. Continuous ensembles

The set of all self-adjoint operators in $\mathcal{H} = \mathbb{C}^n$ has a natural structure of a real space $\mathbb{R}^{2n}$, in which the set of all density matrices is a hypersurface, which is the zero surface $T = 0$ of the affine functional $T = \mathrm{TrX} - 1$.

Let $\mathcal{H} = \mathbb{C}^n$ be an n-dimensional Hermitian space, let ρ be a density matrix in $\mathcal{H}$. We would like to represent the state whose density operator is ρ by an ensemble of pure states. We would like this ensemble to be continuous with the probability density expressed by a function $\mu(\phi)$ where ϕ ranges over all unit vectors in $\mathcal{H}$.
Technical remark. Pure states form a projective space rather than the unit sphere in $\mathcal{H}$. On the other hand, one may integrate over any probabilistic space. Usually distributions of pure states over the spectrum of observables are studied, sometimes probability distributions on the projective spaces are considered [5]. In this paper for technical reasons I prefer to represent ensembles of pure states by measures on unit vectors in $\mathcal{H}$. We use the Umegaki measure on $\mathbb{C}B_n$— the uniform measure with respect to the action of $U(n)$ normalized so that $\int_{\mathbb{C}B_n} \mathrm{d}\psi = 1$.
Effective definition. The density operator of a continuous ensemble associated with the measure $\mu(\phi)$ on the set $\mathbb{C}B_n$ of unit vectors in $\mathcal{H}$ is calculated as the following (matrix) integral

$$\rho = \int\limits_{\phi \in \mathbb{C}B_n} \mu(\phi)\ |\phi\rangle\langle\phi|\ \mathrm{d}\psi \tag{8}$$

where $|\phi\rangle\langle\phi|$ is the projector onto the vector $\langle\phi|$ and $\mathrm{d}\psi$ is the above mentioned normalized measure on $\mathbb{C}B_n$:

$$\int\limits_{\phi\in\mathbb{C}B_n} \mathrm{d}\psi \;=\; 1 \tag{9}$$

Effectively, the operator integral ρ in (8) can be calculated by its matrix elements. In any fixed basis $\{|\mathbf{e}_i\rangle\}$ in $\mathcal{H}$, each its matrix element $\rho_{ij} = \langle \mathbf{e}_i| \rho |\mathbf{e}_j\rangle$ is the following numerical integral:

$$\rho_{ij} \;=\; \langle \mathbf{e}_i| \rho |\mathbf{e}_j\rangle \;=\; \int\limits_{\phi\in\mathbb{C}B_n} \mu(\phi)\, \langle \mathbf{e}_i|\, \phi\rangle \langle\phi|\, \mathbf{e}_j\rangle \;\mathrm{d}\psi \tag{10}$$

The task of likelihood-based recognition of the initial input coding is solved by introducing a special sort of continuous ensembles: so-called Lazy ensembles [6].

4. Lazy ensembles

Potentially we consider all possible input states, and the result we will find in terms of a distribution on the set of all input states. Our task is to guess (to mostly possible extent) what was the distribution of input states.

Definition of Kullback–Leibler distance. We quantify the state preparation efforts by the difference between the entropy of uniform distribution (that is, our null hypothesis) and the entropy of the ensemble[1] in question. The only obstacle may occur is to define this entropy, let us dwell on it in more detail.

The entropy of a finite distribution $\{p_i\}$ is given by Shannon formula

$$S(\{p_i\}) = -\sum p_i \ln p_i$$

This expression diverges for any continuous distribution: we approximate a continuous distribution $\mu(x)$ by a discrete one $\{p_i\}$, calculate its Shannon entropy, but it tends to infinity as we refine the partition. However, we are always interested in the *difference* between the entropy of the uniform distribution and the distribution $\mu(x)$ rather then the entropy itself. At each approximation step we calculate this difference, and the appropriate limit always exists. To show it (see, e.g. [8] for details), make a partition of the probability space by N sets Δ_i having equal uniform measure. Then the difference E_N between the entropies read:

$$E_N \;=\; \ln N \;-\; \Big(-\sum p_i \ln p_i\Big)$$

where $p_i = \int_{\Delta_i} p(x)dx$. The limit expression $\lim_{N\to\infty} E_N$ is the differential entropy

$$S(\mu) \;=\; \int \mu(x) \ln \mu(x) \mathrm{dx} \tag{11}$$

Remarkably, this is equal to Kullback-Leibler distance [3]

$$S(\mu\|\mu_0) \;=\; \int \mu(x) \ln \frac{\mu(x)}{\mu_0(x)} \mathrm{dx}$$

between the distribution $\mu(x)$ and the uniform distribution $\mu_0(x)$ with constant density, normalize the counting measure dx on the probability space so that $\mu_0 = 1$.

[1]We are speaking here of *mixing entropy* [7] of the ensemble rather than about von Neumann entropy of its density matrix.

This distance is the average likelihood ratio, on which the choice of statistical hypothesis is based. Then, in order to minimize the Type I error we have to choose a hypothesis with the smallest average likelihood ratio.
Maximizing the entropy. The problem reduces to the following. For given density matrix ρ find a continuous ensemble μ having minimal differential entropy (11):

$$S(\mu) \;\to\; \min, \qquad \int |\psi\rangle\langle\psi| \;\mu(\psi)\,\mathrm{d}\psi \;=\; \rho \tag{12}$$

where $\mathrm{d}\psi$ is the unitary invariant measure on pure states normalized to integrate to unity. When there is no constraints in (12), the answer is straightforward—the minimum (equal to zero) is attained on uniform distribution. To solve the problem with constraints, we use the Lagrange multiples method. The appropriate Lagrange function reads:

$$\mathcal{L}(\mu) \;=\; S(\mu) \;-\; \mathrm{Tr}\,\Lambda \left(\int |\psi\rangle\langle\psi| \;\mu(\psi)\,\mathrm{d}\psi \;-\; \rho \right)$$

where the Lagrange multiple Λ is a matrix since the constraints in (12) are of matrix character. Substituting the expression (11) for $S(\mu)$ and making the derivative of $\mathcal{L}$ over μ zero, we get

$$\mu(\psi) \;=\; \frac{\mathrm{e}^{-\,\mathrm{TrB}|\psi\rangle\langle\psi|}}{Z\,(B)} \tag{13}$$

where B is the optimal value of the Lagrange multiple Λ which we derive from the constraint (12) and the normalizing multiple

$$Z(B) \;=\; \int \mathrm{e}^{-\,\mathrm{TrB}|\psi\rangle\langle\psi|}\,\mathrm{d}\psi \tag{14}$$

is the partition function for (13). Substituting he resulting density (13) to the expression (11) for differential entropy we get

$$S \;=\; \mathrm{TrB}\rho \;-\; \ln \mathrm{Z} \tag{15}$$

5. Conclusions

It follows directly from the Holevo bound formula that the classical communication capacity of a quantum channel increases when pure states are used for input coding. In the meantime the idea to represent the input ensemble by minimal number of input states, that is, to make them orthogonal, does not in general increase the efficiency of the channel, some coding schemes are essentially based on non-orthogonal states [1]. The usage in statistics of non-orthogonal, overfilled bases of pure states, that is, using the randomization, may sometimes bring some gain in *identifying* the state of the system. This is an essentially quantum phenomenon as in the classical case randomization only produce problems in state identification.

In this contribution we are dealing with the extreme case of overfilled systems of pure states, namely, in a finite-dimensional space we consider bases of infinitely many, continuously many pure states. The research along this lines was first carried out in [9], where the so-called 'Scrooge' ensembles were introduced as bringing minimal amount of information about the preparation procedure. These ensembles

turn out to be continuous. The closest analogy to them in quantum coding are mixed sources [10]

The relevance of continuous ensembles is also justified when we take into account the fact that even we use discrete ensembles of pure states, in reality, when we are preparing them, we can not completely avoid noises produced by the measurement apparata, that is, the distribution of really prepared pure states again turns out to be continuous. We use continuous ensembles for the purpose of building statistical inferences according to the standard schemes of re-estimation of hypotheses, which looks as follows. We have a null hypothesis, then we acquire some information about the system, and then we pass to a new, concurring hypothesis choosing it in such a way that it should be closest to the null hypothesis. This new hypothesis is put forward according to the concept of maximal likelihood.

The concept of maximal likelihood technically reduces to maximization of the logarithm of the probability of the distribution. The appropriate opposite value is a well-known L.J.Savage entropy, or Kullback-Leibler distance between distributions.

In other words, we use the acquired information about the system with a maximal precaution in order to minimize the Type II error.

In our case the null hypothesis is that all pure states are equiprobable. In this case Kullback-Leibler distance between the new ensemble and the null hypothesis is equal to the differential entropy of the new ensemble. The source of additional information is the average density matrix of the input ensemble.

Optimal ensembles that we obtain are exponential distributions of pure staes, which average to a density matrix ρ. These distribution have a striking similarity with the Gibbs ensembles, which form the basis of statistical physics.

The matrix parameter B here plays a rôle similar to that of the temperature in thermodynamics. Under appropriate normalization the value

$$\mathrm{TrB}\rho$$

Will be equal to the differential entropy of the ensemble. So, we may treat this parameter as the differential entropy of the density matrix.

References

[1] Schumacher B, Westmoreland M, *Relative entropy in quantum information theory*, Quantum Computation and Quantum Information: A Millenium Volume , S.Lomonaco, editor (American Mathematical Society Contemporary Mathematics series, 2001); arXiv:quant-ph/0004045

[2] Edwin Thompson Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press (2003)

[3] Kullback S, *Information theory and statistics*, New York, Dover (1968)

[4] N. Hadjisavvas, The Maximum Entropy Principle as a consequence of the principle of Laplace, *J. Stat. Phys.* **26**, 807–815 (1981)

[5] E. Lehrer, E. Shmaya, *A Subjective Approach to Quantum Probability*; eprint quant-ph/0503066

[6] George Parfionov, Roman Zapatrin, 'Lazy' quantum ensembles, *Journal of Physics A: Mathematical and General*, **39** 10891–10900 (2006), arxiv:quant-ph/0603019

[7] Wehrl A, *General properties of entropy*, Reviews of Modern Physics, **50**, 221–260 (1978)

[8] Stratonovich R L, *Information theory* (in Russian), Moscow, Nauka (1975)

[9] R.Jozsa, D.Robb and W.K.Wootters, A Lower Bound for Accessible Information in Quantum Mechanics, *Physical Review* **A49**, 668-699 (1994)

[10] Garry Bowen, Nilanjana Datta, Quantum oding Theorems for Arbitrary Sources, Channels and Entanglement Resources, arXiv:quant-ph/0610003v1