Information Gain in Quantum Theory

by

Mohammad Faghfoor Maghrebi

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The Faculty of Graduate Studies

(Physics)

The University Of British Columbia

(Vancouver)

July, 2008

Abstract

In this thesis I address the fundamental question that how the information gain is possible in the realm of quantum mechanics where a single measurement alters the state of the system. I study an ensemble of particles in some unknown (but product) state in detail and suggest an optimal way of gaining the maximum information and also quantify the corresponding information exactly. We find a rather novel result which is quite different from other well-known definitions of the information gain in quantum theory.

Table of Contents

Appendices

List of Figures

Acknowledgements

I would like to thank my supervisor Dr. Gordon Semenoff who firstly and warmly welcomed me as a prospective student. While I had learnt a lot from his amazingly wide knowledge on various topics which I had the benefit to collaborate with him, I wish also to thank him for he was patiently open to my ideas and gave me much room to explore and accomplish this project. I would also like to thank Dr. Robert Raussendorf with whom I had some very inspiring and valuable discussions. He helped me a lot to come to a better understanding of some parts of the project and provided me with some valuable references. I would like to thank Rogayeh, my wife, with whom I shared the memories and ups and downs of the last few years specially since we came to Canada. She helped me a lot in preparing my notes and made me look more intelligible. I have been always encouraged by her presence in all my explorations and I loved the way we learnt things together. I would like also to thank Jennifer Godfrey who helped me a lot in the final editing of this letter.

Dedication

To my parents,

To my Mother who has been with me where nobody else was.

To my Father who has seemed extremely patient in every step we went forward while he was extremely concerned.

To their unbounded care and love for us.

Chapter 1

Introduction

Laws of nature should make no distinction between reality and information.

Anton Zeilinger

The emergence of the macroscopic world from quantum mechanics is not well understood. What the wavefunction represents in quantum mechanics is still a matter of ongoing debate. One of the most promising approaches in quantum mechanics is Quantum Information theory which argues that the wavefunction is related to the information in some way [1, 2]. The many-world interpretation of quantum mechanics also regards the wavefunction as the complete representation of many parallel worlds and thus assigns some physical interpretation to the wavefunction[3, 4]. There is also the Bayesian approach [5] which argues that quantum mechanics should be viewed as a Bayesian system in which all statements are regarded as an agent's degree of belief and it has nothing to do with a pre-existing reality. The mere existence of multiple approaches illustrates the bizarre situation.

There have been even suggestions in the past that the wavefunction could be measured under certain circumstances [6] but it has also been argued that there is no ontological sense of the wavefunction beyond its epistemological meaning [7]. That is, the wavefunction is not *real* and thus can not be *measured* or determined [8]. However we are certainly measuring something in the lab. Measuring the quantum state is actually an entire field of study in itself. This usually means that there are many particles available in the same state. The quantum state can be then determined by various methods such as tomography, etc. [9]. Therefore, knowing certain information (in this case, many particles in the same state) about the wavefunction, we can determine more (find the state). As an interesting example in which we can actually determine whether some state is entangled or not, suppose we have two identical systems, say A and B [10]. The state is thus $|\psi\rangle \otimes |\psi\rangle$. Each system has two particles in some arbitrary state (entangled or not)

$|\psi\rangle \in \mathcal{H}_{\{1_s,2_s\}}$ where $s = A, B$ and $\mathcal{H}_{\{1_s,2_s\}}$ is the Hilbert space of the two particles. We can then determine if these two particles are entangled or not by a single measurement. It might seem that this is possible because we have two copies. However, the fact that there are two copies is not relevant because we can think of the two systems as a single system in which the total state (as a product state) $|\psi\rangle_{A+B} = |\psi\rangle_A \otimes |\psi\rangle_B$ is an eigenstate of the exchange operator $P_{A,B}$, which exchanges the two subsystems. Knowing this, we can determine whether 1 and 2 are entangled or not. There is also a nice discussion that considers more carefully some of the assumptions of the original reference[11]. Another perhaps the most famous example is Quantum Information theory which assumes an ensemble $\mathcal{E} = \{\rho_x, p_x\}$ where p_x is the probability that state ρ_x is sent on a communication channel. The problem is then to find the maximum information by performing some measurement on the states received from the first party. Von Neumann entropy and the accessible information are central in this framework.

In all these examples, we need some *a priori* knowledge to get more information from the system. If there is no *a priori* knowledge, it seems we can not extract any information whatsoever about the system. However, as an example, there is a huge knowledge of distant galaxies and their extensive properties. How is this information gained? What is the *a priori* knowledge? And how much is the information gain?

A preliminary answer can be the following. Let's think of a mechanism of continuous measurement which projects the state of the system onto a special basis[12]. Knowing this basis, we can find the state of the system (by a projective measurement in this basis). For example, let's suppose that we are living in some environment that the particles are continuously measured in the position basis[1]. We can determine the state by measuring the position.

However this process (of continuous measurement) works only if there is a special dynamics. For example, the dynamics should localize the particles. In general, There might be no such dynamics. As an example, consider the polarization of a photon (or a beam of photons). There is no preferred direction in the space along which the state must collapse. The rotation symmetry of the world might lead to a collapse of the photon state in any direction. So while this answer could be partially true it certainly does not

[1]We are assuming that the measurement frequency is slow enough so the wavefunction would actually evolve (as opposed to Zeno Paradox case) but fast enough with respect to some macroscopic time scale [12].

tell the whole story.

A simple case of interest is when there are many particles almost in the same state. For example, radiation from a distant galaxy might contain subsystems each consisting of many particles in an (almost) identical state. Can we justify that they are truly identical? The answer is affirmative. For instance, in his book, Asher Peres suggests to divide the particles into many subgroups of large numbers and repeat a series of measurements on these subgroups[13]. If the ensemble is truly identical, we must have obtained consistent results. Note that it is assumed that the overall state is a product state of the particles.

This can partially answer the problem we posed, however it is not quite clear what the answer is if not all the particles are in the same state or why one may assume that the overall state is a product state. Also the amount of information acquired is not quantified and it is not obvious if this is the most efficient way to maximize the information gain.

In this letter, we generalize this to the case where the particles might be in non-identical states and find the maximum information gain in an efficient way which doesn't disturb the system to high precision. We then find the amount of the information gain which, to some surprise, does not follow any of the usual definitions of the information gain in quantum mechanics.

In order to achieve the answer, we must re-examine some ideas about the emergence of the probability statements in quantum mechanics. The definition and the derivation of the probabilities in quantum mechanics is also unclear. There is a question as to whether we must derive the probabilities from first principles or if we should accept them as part of the principles of quantum mechanics. There have been various attempts to derive the probability laws [3, 14–16]. Part of the problem lies in the definition of the probability itself. Is the probability interpreted as a frequency law [14, 15](how often some state would result) or is it merely a decision-making rationale [16]? We will argue that at least some notion of the probability is actually derived from first principles in quantum mechanics. The result is not novel even though the derivation may be. Once we have the necessary terms to express the probability, we can generalize that to something which gives us the technology to find the information gain. This generalization deals with an ensemble of not-identical states as opposed to the more conventional one in which all particles are assumed to be in the same state.

The structure of this paper is as follows. In section 2.1, we derive the Born rule of probability. We generalize the Born rule to non-identical states

in section 2.2. In sections 3.1 and 3.2 we find the best information gain and derive its upper bound. We then derive the information gain rigorously in section 3.3. Finally, in the last chapter, we make some concluding remarks on the information gain in quantum mechanics and the role of the measurement.

Chapter 2

Probability

2.1 Derivation of Born Probability

In this section we present a simple proof of the Born rule of Probability. We demonstrate this for a two-level system. The arguments may be generalized immediately.

Suppose we have N systems in an identical state where N is assumed to be large. The quantum state is then

$$|\psi\rangle \otimes |\psi\rangle \otimes \cdots \otimes |\psi\rangle = |\psi\rangle^N .$$

Let's consider the following operator

$$\hat{P} = \sum_{\forall\psi\in\text{Hilbert space}} f(\psi)\ |\psi\rangle^N \langle\psi|^N .$$

In the limit of $N \to \infty$, $|\psi\rangle^N$ is an eigenstate of this operator. The reason is that if $|\psi\rangle \neq |\psi'\rangle$ then $|\langle\psi|\psi'\rangle| < 1$ and $|\langle\psi|\psi'\rangle|^N \to 0$ as $N \to \infty$.

Assuming $|\psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle$, we find

$$\begin{aligned} |\psi\rangle^N \langle\psi|^N &= (\alpha|\uparrow\rangle + \beta|\downarrow\rangle)^N (\alpha^* \langle\uparrow| + \beta^* \langle\downarrow|)^N \\ &= \sum_{n,m,\text{ all orderings}} \alpha^n \beta^{N-n} |\uparrow\uparrow\downarrow \ldots\rangle\, \alpha^{*m} \beta^{*N-m} \langle\uparrow\downarrow\uparrow \ldots| . \end{aligned}$$

We then project this operator onto a subspace in which n particles point upward and the rest $N-n$ point downward. Let's call this projector $\hat{P}_{n,N-n}$.

$$\begin{aligned} \text{Tr}(\hat{P}_{n,N-n} |\psi\rangle^N \langle\psi|^N) &= \sum_{n,\text{ all orderings}} (|\alpha|^2)^n (|\beta|^2)^{N-n} \text{Tr}(\underbrace{|\uparrow\downarrow\uparrow \ldots\rangle}_{n's\uparrow, N-n's\downarrow} \langle\uparrow\downarrow\uparrow \ldots|) \\ &= \binom{N}{n} (|\alpha|^2)^n (|\beta|^2)^{N-n}. \end{aligned}$$

This is a very localized function of n. The maximum of this function occurs when $n_*/N = |\alpha|^2$ and $(N - n_*)/N = |\beta|^2$ for $N \to \infty$. The standard deviation in n is $\delta n \sim \sqrt{N}$. Defining

$$\hat{P}_p^\epsilon = \sum_{(p-\epsilon)N \leq n \leq (p+\epsilon)N} \hat{P}_{n,N-n},$$

we find that for sufficiently large N the state $|\psi\rangle^N \langle\psi|^N$ is completely in the subspace given by the projection operator $P^\epsilon_{|\alpha|^2}$, where ϵ could be taken as small as desired for $N \to \infty$, that is

$$\lim_{N\to\infty,\epsilon\to 0} \mathrm{Tr}(\hat{P}_p^\epsilon |\psi\rangle^N \langle\psi|^N) = \begin{cases} 0, & \text{if } p \neq |\alpha|^2 \\ 1, & \text{if } p = |\alpha|^2. \end{cases}$$

Now suppose that we perform some measurement on the individual spin 1/2 particles. We can measure the spin along the z axis of each particle. The measurement yields $|\uparrow\rangle|\uparrow\rangle|\downarrow\rangle\ldots$, for example. The last equation would then imply that

$$\lim_{N\to\infty,\epsilon\to 0} (\langle\uparrow|\langle\uparrow|\langle\downarrow|\ldots)|\psi\rangle^N = 0 \quad \text{if the number of } \uparrow\text{'s in the outcome} \neq N(|\alpha|^2 \pm \epsilon)$$

Therefore, some version of the Born rule is derived here. Our only assumption was that the measurement outcome of some wavefunction will never be some state which has no overlap with the wavefunction.

To interpret these results as probability, we must make some justifications concerning the standard deviation of the results. Let's review the Central Limit Theorem[2]. Assume X_1, X_2, X_3, ... are a set of N independent and identically distributed random variables having mean value μ and variance σ^2. The Central Limit Theorem then implies that as the sample size N increases, the distribution of the sample average approaches the normal distribution with a mean μ and variance σ^2/N. In our problem

$$E(X) = p.0 + (1-p).1 = 1 - p, \quad E(X^2) = p.0^2 + (1-p).1^2 = 1 - p,$$

so

$$\sigma^2 = E(X^2) - E(X)^2 = p(1-p) = |\alpha|^2|\beta|^2.$$

[2]See Appendix C.

Let's see if we would get the same result through our earlier arguments. Defining

$$X_n = \binom{N}{n}(|\alpha|^2)^n(|\beta|^2)^{N-n},$$

we find

$$X_n \approx \frac{1}{\sqrt{\pi}|\alpha||\beta|\sqrt{N}}\exp(-\frac{1}{2}\frac{N(p-p_*)^2}{|\alpha|^2|\beta|^2})$$

where $p_* = |\alpha|^2$. So the standard deviation is exactly what it should be.

We can also find the error *probability* in the case of finite N. We must sum up all X_n with $|n - n_*| \geq N\epsilon$.

$$P_{error} \;\; = \;\; \int_{|n-n_*|\geq N\epsilon} dn \;\; X_n = \frac{2}{\sqrt{\pi}}\int_{\frac{\sqrt{N}\epsilon}{|\alpha||\beta|}}^{\infty} d\tilde{n} \;\; \exp(-\frac{1}{2}\tilde{n}^2)$$

so

$$P_{error} < \frac{2}{\sqrt{\pi}}\frac{|\alpha||\beta|}{\sqrt{N}\epsilon}\exp(-\frac{1}{2}\frac{N\epsilon^2}{|\alpha|^2|\beta|^2}) \tag{2.1}$$

where we have used $\int_{x_0}^{\infty} dx \;\; \exp(-\frac{1}{2}x^2) = \int_{\frac{1}{2}x_0^2}^{\infty} dy \;\; \frac{1}{\sqrt{2y}}exp(-y) < \frac{1}{x_0}exp(-\frac{1}{2}x_0^2)$.

In order to get a good approximation, we must have[3]

$$\epsilon \gg \frac{1}{\sqrt{N}}|\alpha||\beta|.$$

Now let's partition the N-particle unity as follows

$$\mathbf{1} = \sum_p \hat{P}_p^{\epsilon}$$

where we are summing over all p in $\{0, \ldots, \epsilon i, \epsilon(i+1), \ldots, 1\}$. Consider a projective measurement defining by $\{P_p^{\epsilon}\}$. If the result of the measurement turns out to be p, we find

$$\Big| \;\; |\alpha|^2 - p\Big| \leq \epsilon$$

where the error probability is as given before and could be taken as small as desired for $N \to \infty$.

[3]Note that this relation doesn't mean that ϵ could be zero when $\alpha = 0$ since the best we can get is to find $|\alpha|$ and $|\beta|$ within some approximation. There is always some uncertainty ϵ. In the extreme case, we find $|\alpha|^2 \leq \epsilon$. Plugging back to the last equation, this gives $\epsilon \gg \frac{1}{N}$.

The whole procedure is equally applicable to the mixed states. The proof is exactly the same. We can find the overlap of $\rho^{\otimes N}$ with $P^z_{n,N-n}$ which projects n particles onto $|\uparrow\rangle$ and projects the rest onto $|\downarrow\rangle$

$$\mathrm{Tr}(\rho^{\otimes N} P^z_{n,N-n}) = \mathrm{Tr}(\rho_z^{\otimes N} P^z_{n,N-n}).$$

Therefore we can practically substitute ρ by the reduced density operator ρ_z which is defined as

$$\rho_z = \begin{pmatrix} p_1 & 0 \\ 0 & p_2 \end{pmatrix}.$$

So everything also applies to a mixed state if we substitute p_1 for $|\alpha|^2$ and p_2 for $|\beta|^2$. Then the probability is found by

$$Prob(i) = \mathrm{Tr}(\rho \hat{P}_i)$$

where i is a certain outcome of an experiment and $\hat{P}_i$ is the corresponding operator.

There have been some proofs of the Born rule in the frequency interpretation [14, 15] in the literature. However it has been objected to by some physicists. Wallace [17], for example, argues that infinity never occurs in real life or in any finite-event scenario. He points out that in any statistically finite-numbered experiment we can not neglect the tail of the distribution merely because it is very small. He also argues that the most natural framework to derive the probability is the many-worlds interpretation of quantum mechanics and some improvements of what has already been developed by Duetsch [16] will do that.

2.2 Ensemble of Particles in a Product State

The statistics of the outcomes of a quantum measurement obey a probabilistic pattern. So we can use the full strength of the probability theory.

We wish to generalize the previous results to an ensemble of particles in arbitrary (but product) state. After all, we are interested in knowing how and by how much we can gain information from the state of a macroscopic system where all the particles may not be in the same state.

The Lyapunov's Central Limit Theorem will be proved to be useful. Let X_n be a sequence of independent random variables. Suppose that the third central moments

$$r_n^3 := \mathbf{E}[\big|X_n - \mu_n\big|^3]$$

are finite and satisfy the Lyapunov condition[4]

$$\lim_{N\to\infty} \frac{\left(\sum_{n=1}^{N} r_n^3\right)^{1/3}}{\left(\sum_{n=1}^{N} \sigma_n^2\right)^{1/2}} = 0.$$

Let the random variable $S_N := X_1 + \cdots + X_N$ denote the sum of the random variables X_n. For "large" N, S_N is normally distributed[5] with the expected value

$$\mathbf{E}[S_N] \approx \sum_{n=1}^{N} \mathbf{E}[X_n]$$

and the variance

$$\mathrm{Var}[S_N] \approx \sum_{n=1}^{N} \mathrm{Var}[X_n].$$

Let's define the total variance as $\sum_{n=1}^{N} \mathrm{Var}[X_n] = N\sigma^2$ where σ is the average of the variance. In our problem, however, there are only two possible outcomes which can be chosen as 0, 1 (that stand for $|\uparrow\rangle$ and $|\downarrow\rangle$ for example). Therefore we have the following identity ($|X_n - \mu_n| \leq 1$)

$$r_n^3 := \mathbf{E}[\big|X_n - \mu_n|^3] \leq \mathbf{E}[\big|X_n - \mu_n|^2] =: \sigma_n^2$$

or $\sum_{n=1}^{N} r_n^3 \leq N\sigma^2$. Then the ratio of the third moment to the second one is

$$\lim_{N\to\infty} \frac{\left(\sum_{n=1}^{N} r_n^3\right)^{1/3}}{\left(\sum_{n=1}^{N} \sigma_n^2\right)^{1/2}} \leq \frac{1}{(N\sigma)^{1/6}}.$$

The Lyapunov condition is then satisfied for $\sigma \gg 1/N$. In other words, this means that we can not find the expected value (i.e. the average of the random variables) with a resolution beyond $1/N$.

The requirements of the theorem is satisfied in our problem. The total variance is then the sum of all the variances

$$\sigma_{TOT}^2 = \sum_{n=1}^{N} \mathrm{Var}[X_n] = N\sigma^2$$

[4]see Appendix D

[5]see Appendix C

and the probability of error is given by[6]

$$P_{error} \leq \frac{\sigma^2}{N\epsilon^2}$$

where ϵ is the resolution. As a result, ϵ must be chosen such that[7]

$$\epsilon > \sigma/\sqrt{N}.$$

In conclusion, we find that, up to a negligible error, the average of the density operators of the states can be found.

We can also deduce the same result in the more familiar operators' language. Consider the state

$$|\psi\rangle = \alpha\,|\uparrow\rangle + \beta\,|\downarrow\rangle\,.$$

For a large ensemble of identical states, we have already showed that $|\psi\rangle^{\otimes N}$ would be *almost* an eigenstate of $\hat{P}_p$ (we dropped the superscript ϵ) with the eingenvalue 1 if $p \approx |\alpha|^2$ and 0 otherwise. The N-particle state can be grouped as

$$|\Psi\rangle = |\psi_1\rangle^{n_1} \otimes |\psi_2\rangle^{n_2} \otimes \cdots \otimes |\psi_M\rangle^{n_M}$$

where n_1 particles are almost in the same state $|\psi_1\rangle$, n_2 particles are almost in the same state $|\psi_2\rangle$ and so forth.

It is then true that $\hat{P}_{p_i}^{(n_i)}\,|\psi_i\rangle^{n_i} = |\psi_i\rangle^{n_i}$ if $p_i \approx |\alpha_i|^2$ and it's 0 otherwise. The superscript n_i represents the number of particles in the same state.

Now consider the operator $\hat{P}_p^{(N)}$ which acts on the N-particle-space. We can expand this operator as

$$\hat{P}_{p'}^{(N)} = \sum_{p'_1,p'_2,\ldots,p'_M} \hat{P}_{p'_1}^{(n_1)} \otimes \cdots \otimes \hat{P}_{p'_M}^{(n_M)}$$

where the summation is constrained by $\frac{1}{N}(n_1 p'_1 + n_2 p'_2 + \cdots + n_M p'_M) = p'$.

Applying this operator to $|\Psi\rangle$, only those terms in which $p'_1 = p_1, \ldots, p'_M = p_M$ will contribute. In conclusion, $|\Psi\rangle$ is *almost* an eigenstate of the operator

[6] see Appendix B

[7] In the extreme case where the distribution is very close either to 1 or to 0, we have $\epsilon > 1/N$.

$\hat{P}_p^N$ with the eigenvalue 1 if $p \approx \frac{1}{N}(|\alpha_1|^2 n_1 + |\alpha_2|^2 n_2 + \cdots + |\alpha_M|^2 n_M)$ and 0 otherwise. We can also write the last condition in a more suggestive way:

$$p \approx \frac{1}{N} \sum_{i=1..N} |\alpha_i|^2 \tag{2.2}$$

So for a large number, we can find the mean probability distribution to high precision.

Theses results will be readily generalized to the product of the mixed states. Also they will be exact for $N \to \infty$. That is, we find

$$\lim_{N\to\infty,\epsilon\to 0} \mathrm{Tr}(\hat{P}_p^{(N),\epsilon} \rho_1 \otimes \rho_2 \otimes \rho_3 \otimes \dots) = \begin{cases} 0, & \text{if } p \neq \bar{p} \\ 1, & \text{if } p = \bar{p} \end{cases}$$

where $\bar{p}$ is defined as

$$\bar{p} = \frac{1}{N} \sum_i \langle \uparrow | \rho_i | \uparrow \rangle.$$

The interesting feature here is that the mean probability ($\bar{p}$) can be determined by a collective measurement that doesn't alter the state (because it is an eigenstate of the measuring operators). So we can perform more measurements (a spin 1/2, for example, can be also measured along the x and y axes in addition to the previously-measured spin z). In this way we can find the average of the *density operators* of individual particles[8]

$$\bar{\rho} = \frac{1}{N} \sum_{1..N} \rho_i. \tag{2.3}$$

[8]This is also in agreement with [18] which argues that in order to perform the optimal measurement we should consider the whole ensemble as a single system rather than a sum of its components.

Chapter 3

Information Gain

3.1 Best Information Gain

So far we have proved that giving an ensemble of particles in a product state, we can find the average of the density operators of individual particles. Since we are interested in the best information gain, we should ask if there is any information available to the observer beyond this. Can we, for example, find the "n-th moment" (with a slight abuse of the word "moment") of the density operators of the particles defined as

$$\mu_n = \frac{1}{N} \sum_{1..N} \rho_i^n$$

(ρ_i is the density operator of the i-th particle)?

We argue that finding anything beyond the first moment (i.e. the average of the density operators) would contradict the no-signalling theorem. The following is a thought experiment which demonstrates this point.

Suppose we have a large ensemble of particles which are entangled in pairs. The two particles in each pair are taken far apart. We then have two groups of particles in which each particle from one group is entangled to another from the other group.

We would like to perform some measurement on the second group of particles. Consider some generalized measurement, the POVM $\{M_m\}$ which is acting on one-particle states in the second group. They would satisfy

$$\sum_m M_m^\dagger M_m = \mathbf{1}$$

where $\mathbf{1}$ is the unity operator in the one-particle Hilbert space. Suppose that some pair is in the state $|\Psi_{12}\rangle$, where the subindexes 1, 2 refer to the two particles. For a large ensemble of the particles, we can assume that there are very many pairs almost in the same state[9]. Then as we perform

[9]If this is not the case for a subset of particles, we can safely neglect them.

the measurement, a fraction $\|\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle\|^2$ of the pairs in this state will collapse to

$$\frac{\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle}{\|\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle\|}.$$

As we studied in detail, the average of the density operators can be measured. So after that the measurement has been performed on the second group, the local observer at the first group decides to measure the average density operator and finds

$$\begin{aligned}&\frac{1}{N}\sum_{\text{states}} n\sum_m \|\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle\|^2 \mathrm{Tr}_2\left(\frac{\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle}{\|\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle\|}\frac{\left\langle\Psi_{12}\right|\mathbf{1}\otimes M_m^\dagger}{\|\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle\|}\right)\\ =\ &\frac{1}{N}\sum_{\text{states}} n\mathrm{Tr}_2\left(\left|\Psi_{12}\right\rangle\left\langle\Psi_{12}\right|\mathbf{1}\otimes\sum_m M_m^\dagger M_m\right)=\frac{1}{N}\sum_{\text{states}} n\rho\end{aligned}$$

where $\rho=\mathrm{Tr}_2\left(\left|\Psi_{12}\right\rangle\left\langle\Psi_{12}\right|\right)$ and n is the number of particles which were initially almost in the state $\left|\Psi_{12}\right\rangle$. Note that we are also summing over all the states.

However, this is also the average of the density operators right before any measurement was performed. So it is independent of what specific POVM has been chosen; any other POVM would result in the same average density operator. There is no way to determine what POVM was used, given only the average of the density operator in the first group. That is there is no way for signalling at a distance to occur.

On the contrary, any knowledge beyond the average of the density operators would convey some information about which POVM has been chosen and it would then lead to superluminal communication. As an example, we can find the second moment of the reduced density operator of the first group (after the measurement has been done on the second group). It is

$$\sum_{m,\ \text{states}} n\mathrm{Tr}_2\left(\left|\Psi_{12}\right\rangle\left\langle\Psi_{12}\right|\frac{\mathbf{1}\otimes M_m^\dagger M_m}{\|\mathbf{1}\otimes M_m\left|\Psi_{12}\right\rangle\|}\right)^2\neq(\sum_{\text{states}} n\rho^2)/N$$

which vividly depends on the choice of the POVM.

This can be proved in general as follows. Assume that initially all the particles are in the same state $\left|\Psi_{12}\right\rangle=\left|\Psi\right\rangle$. After performing the measurement a fraction $\|\mathbf{1}\otimes M_m\left|\Psi\right\rangle\|^2$ of the pairs in this state will collapse to $\frac{\mathbf{1}\otimes M_m\left|\Psi\right\rangle}{\|\mathbf{1}\otimes M_m\left|\Psi\right\rangle\|}$.

We can then trace over the second group of the particles. The reduced state is then

$$\frac{\mathrm{Tr}_2\left(|\Psi\rangle\langle\Psi|\,\mathbf{1}\otimes M_m^\dagger M_m\right)}{\mathrm{Tr}_{1,2}\left(|\Psi\rangle\langle\Psi|\,\mathbf{1}\otimes M_m^\dagger M_m\right)}$$

Now we look for a function of the reduced density operators which is indifferent to the POVM chosen. Note that a special POVM might be $\{\mathbf{1}\}$ which does not do anything. The problem is then to find the most general function with the property

$$f(\rho_1,\rho_2,\ldots,\rho_N) = f(\rho,\rho,\ldots,\rho)$$

where $\rho = \mathrm{Tr}_2(|\Psi\rangle\langle\Psi|)$ and ρ_i's are the reduced density states as defined in the above and they are constrained as

$$\frac{1}{N}\sum \rho_i = \rho. \tag{3.1}$$

It follows that $f(\rho_1,\rho_2,\ldots,\rho_N)$ must be only a function of the average of its arguments.

$$f(\rho_1,\rho_2,\ldots,\rho_N) = F(\rho = \mathbf{E}(\rho_i)).$$

That is any such function only carries information about the average density operator.

To give an honest proof, we must show that $\{\rho_i\}$ are only constrained by 3.1. That is they are not constrained any further by the measurement, i.e. the POVM $\{M_m\}$. In other words, we must show for any $\{\rho_i\}$ constrained to 3.1, there is always some POVM which results in the same set of states.

We choose another alternative here as follows. As we have said, a fraction $\|\mathbf{1}\otimes M_m|\Psi\rangle\|^2$ of the pairs will collapse to

$$\frac{\mathrm{Tr}_2\left(|\Psi\rangle\langle\Psi|\,\mathbf{1}\otimes M_m^\dagger M_m\right)}{\mathrm{Tr}_{1,2}\left(|\Psi\rangle\langle\Psi|\,\mathbf{1}\otimes M_m^\dagger M_m\right)}.$$

Let's introduce the (non-normalized) operators $\tilde{\rho}_m$ as

$$\tilde{\rho}_m = \mathrm{Tr}_2\left(|\Psi\rangle\langle\Psi|\,\mathbf{1}\otimes M_m^\dagger M_m\right).$$

Normalizing this operator, we find the reduced state and its norm also gives us the fraction of the particles reduced in this state. The problem can be then reformulated to find a function as

$$f(\tilde{\rho}_1,\tilde{\rho}_2,\ldots,\tilde{\rho}_m,\ldots)$$

which *does not* depend on the POVM M_m's. Since M_m's are only constrained by 3.1, we find[10]

$$\frac{\partial f}{\partial (M_m)_{ij}} - \lambda_k (M_m)^*_{ik} = 0$$

where λ_k's are constant. It follows that

$$\frac{\partial f}{\partial (M_m^\dagger M_m)_{ij}} = const,$$

as we wished to show. Note that this relation would not hold if f is defined as the "n-th moment" of the density operator for any $n > 1$. The proof is then complete.

Note that although we can only find the average density operator, the knowledge of the density operator alone constrains the higher moments because it is a positive trace class operator. The extreme scenario is when the average density operator is a one-dimensional projection operator $\rho = \rho^2$. In this case, the knowledge of the average density operator will also determine all the higher moments.

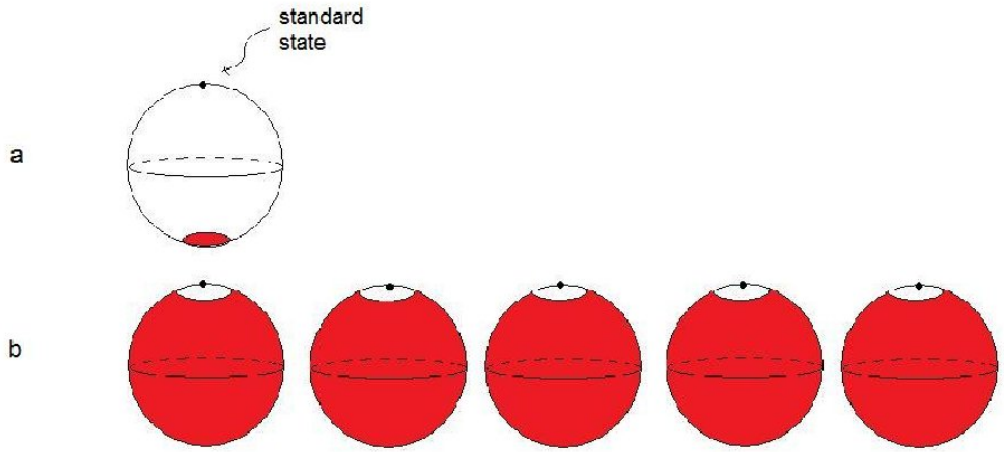
3.2 A Bound on the Information Gain

Before discussing the information gain in full generality, it is interesting to explore whether there is any theoretical limit on the information gain of a system of N particles in a two-level system (where we also assume that the particles are in a product state). By general considerations, we can find a bound on the maximum information. This bound comes from the fact that the non-orthogonal states can not be discriminated perfectly [1]. The discrimination of two states is related to their overlap, i.e. how orthogonal they are. We can discriminate two non-orthogonal states up to some error comparable to their overlap. As long as there is a tiny overlap, we can discriminate the states almost reliably. To get some insight, let's define the two states to be "ϵ-distinguishable" if (the absolute value of) their inner product is less than ϵ. Let's consider a single spin 1/2 particle in some state $|\uparrow\rangle$. The fraction of the Hilbert space (in this case the Bloch sphere) that is ϵ-distinguishable from this state is something of the order of ϵ: the tiny

[10] Here we exploit the advantage of using the generalized measurement POVM instead of the projective measurement. M_m's can be chosen arbitrarily subject only to 3.1.

area near the south pole of the Bloch sphere. However, for a large number of particles, this could be dramatically larger. Figure (3.1) illustrates the point. Each sphere is a Bloch sphere. The big dots on top of the spheres represent the standard state where in the first case is $|\uparrow\rangle$ and in the second case, it is $|\uparrow\uparrow\uparrow\ldots\rangle$. The dark area represents the portion of the Hilbert space which is ϵ-distinguishable from the standard state.

Figure 3.1: The ϵ-distinguishable region for the Hilbert space of a) a single particle, b) multi-particles.



We can very easily calculate the minimum number of the states which could be distinguished from the rest of the Hilbert space reliably[11]. Take a standard state such as

$$|\uparrow\uparrow\ldots\uparrow\rangle$$

and another one of the form

$$|\nearrow_{\hat{n}}\nearrow_{\hat{n}}\cdots\nearrow_{\hat{n}}\rangle$$

where the subindexes $\hat{n}$ is the direction of the spin. In order to have a small error probability, the overlap should be small

$$\langle\ \ |\ \ \rangle = |\langle\uparrow|\nearrow_{\hat{n}}\rangle|^N = \cos^N(\theta/2).$$

[11]A similar line of thought is also taken in [19] where coarse-graining is considered necessary in order to find some sort of reality in quantum mechanics.

For the inner product to be small, we must have $N\theta^2 \geq 1$ or

$$\theta \geq 1/\sqrt{N}.$$

Any state such as

$$|\nearrow_{\hat{n}_1}\nearrow_{\hat{n}_2}\cdots\nearrow_{\hat{n}_N}\rangle$$

where all $\hat{n}$'s are confined in the small region of $\theta < 1/\sqrt{N}$, will not be distinguishable from the standard state while any other state would be "almost" distinguishable. The corresponding *volume* (of the Hilbert space that can not be distinguished from the standard state) is then (we have normalized the total volume to 1)

$$\mathcal{N} \sim (\theta^2)^N \sim (1/N)^N.$$

So from this point of view, we find the bound[12]

$$I_{max} \leq N \log N. \tag{3.2}$$

Note that this bound was found merely by imposing the indistinguishability theorem. There is no reason to expect to gain this much information and most of the times one will not. However, we will show that this bound may be saturated.

3.3 Information Gain

In this section, we finally go back to our initial motive and find the information gain. The Shannon information is not appropriate for our purpose for the following reason. The Shannon information (and its immediate quantum mechanical generalization, Accessible Information) requires the *a priori* probability p_x that some state ρ_x may occur. However, there is no such *a priori* probabilities in the problem that we are interested in this letter. We will argue more extensively in the next chapter why this is the case.

We will seek for a definition which is applicable to any sort of *a priori* information about the system. We will state our definition in the physicist's language of the measurement and the state but it is meant to be general.

Consider a system which is going to be measured. Prior to the measurement, the system could be in any of $\mathcal{N}_{total}$ (*likely*) states (the notion of the state doesn't necessarily refer to the quantum state). After the measurement

[12] We will define the information rigourously in the next section.

has been performed, we find that the system was initially *most likely* in some subspace (of the total space) that contains only $\mathcal{N}_{measured}$ states. Note that "*likely*" means the error due to neglecting the rest of the states is small. The information is then defined as

$$I_{gain} = \log(\mathcal{N}_{total}) - \log(\mathcal{N}_{measured}). \tag{3.3}$$

This formula could be interpreted in the following way. When there are $\mathcal{N}$ different possible outcomes and we don't know which one is true, the uncertainty (entropy) would be $\log(\mathcal{N})$. After the measurement, the number of possibilities is reduced. The information gain is defined as the decrease of the uncertainty (entropy) before and after the measurement.

This definition of the information gain is additive in the following sense. The information in two independent (and non-entangled) systems is the sum of the information in each system

$$I_{total} = I_1 + I_2.$$

The crucial feature of this notion of the information is that it's not limited to the *a priori* knowledge that is usually assumed in the application of the Shannon information, i.e. we have not assumed that the states are drawn from an ensemble $\mathcal{E} = \{\rho_x, p_x\}$. It's easy to see that in the case of a (classical) ensemble $\mathcal{E} = \{x, p_x\}$, we would recover the Shannon information. This will be shown in Appendix A.

Note that the information gain as defined here seems to depend on the error that can be allowed. So we must also represent the error probability in the definition of the information $I_{gain}(P_{error})$. A sensible definition of the information gain had better not depend too sensitively on the error. We will come back to this point later.

Now we are in a position to find the information gain. Suppose we are given a large number of spin 1/2 particles which are in a product state. We can measure them along the z axis. We assume further that $n = Np$ of the spins will end in state $|\uparrow\rangle$ after the measurement. According to the (generalized) rule of large numbers, we can find

$$\bar{p} = \mathrm{Tr}(\bar{\rho}|\uparrow\rangle\langle\uparrow|)$$

where $\bar{\rho}$ is the average of the density operator of the individual particles (as defined earlier in equation (2.3)). To be more accurate, we find that

$$\begin{cases} |p - \bar{p}| < \epsilon \\ P_{error} < \frac{\sigma^2}{N\epsilon^2}. \end{cases} \tag{3.4}$$

In order to define the number of states we must *discretize* the Hilbert space. This procedure must preserve the natural symmetries of the Hilbert space. For a two-level system this is quite easy because there is a geometrical picture of the Hilbert space, the Bloch Sphere. The natural measure on the Bloch Sphere is $d\cos\theta d\phi$. This can be rewritten as

$$d\cos\theta d\phi \sim d\cos^2(\theta/2)d\phi \sim dpd\phi$$

where p is the probability of finding the particle in the state $|\uparrow\rangle$. So the appropriate measure is dp (measuring the spin z the particles would reveal no information about ϕ). The number of the states in some interval is then given by

$$\frac{1}{\delta}\int dp,$$

where δ represents the *size* of the discretized volume. Then the total number of states of N particles is given by

$$\mathcal{N}_{total} = (\frac{1}{\delta})^N.$$

We will also find $\mathcal{N}_{measured}$ as

$$\mathcal{N}_{measured} = \frac{1}{\delta}\int' dp_1 \;\; \frac{1}{\delta}\int' dp_2 \ldots \frac{1}{\delta}\int' dp_N$$

where the prime indicates that the integral is constrained by $|p-\bar{p}| < \epsilon$. Note that $\bar{p} = \frac{1}{N}\sum p_i$ in which p_i is the probability of finding the i-th particle in the state $|\uparrow\rangle$ and p is defined as $n_\uparrow = pN$ ($n_\uparrow$ is the number of spins we find in $|\uparrow\rangle$ after the measurement). So

$$\mathcal{N}_{measured} = (\frac{1}{\delta})^N \int_{|\frac{1}{N}\sum p_i - p|<\epsilon} dp_1 dp_2 \ldots dp_N. \tag{3.5}$$

Before trying to evaluate this integral let's get a little bit of insight into its geometrical meaning. The probability space is actually an N-dimensional (super)cube: $0 \le p_i \le 1$. We want to find the volume of the portion of the cube which lies between two parallel (super)planes defined as

$$p_1 + p_2 + \cdots + p_N = N(p \pm \epsilon).$$

Let's denote the volume which is bound by

$$p_1 + p_2 + \cdots + p_N = n$$

and

$$p_1 + p_2 + \cdots + p_N = n + 1$$

by $\mathcal{V}_n$. The number of the *measured* states is

$$\mathcal{N}_{measured} = (\frac{1}{\delta})^N \sum_{N(p-\epsilon)\leq n\leq N(p+\epsilon)} \mathcal{V}_n.$$

Since the distance between the two parallel surfaces is $1/\sqrt{N}$ the volume is approximately

$$\mathcal{V}_n = \frac{1}{\sqrt{N}}\mathcal{A}_n \tag{3.6}$$

where $\mathcal{A}_n$ is the $N-1$ dimensional area of the plane (which is defined as the intersection of the super-cube and the super-plane $p_1 + \cdots + p_N = n$). The area is a well-known series expansion [20]:

$$\mathcal{A}_n = \frac{\sqrt{N}}{(N-1)!}\sum_{k=0}^{[n]}(-1)^k\binom{N}{k}(n-k)^{N-1}. \tag{3.7}$$

This can also be written as an integral

$$\mathcal{A}_{n=Np} = \frac{2\sqrt{N}}{\pi}\int_0^\infty du\left(\frac{\sin u}{u}\right)^N \cos(2N(p-1/2)u). \tag{3.8}$$

However this a difficult integral to evaluate even numerically and there is no general solution in the literature to the best of our knowledge. Yet we can consider some special cases here.

For small $|p-1/2|$, we find

$$\mathcal{A}_{Np} = \sqrt{\frac{6}{\pi}}\exp(-6N(p-1/2)^2). \tag{3.9}$$

The standard deviation around $p = 1/2$ is then $1/\sqrt{N}$. But our error estimation (3.4) also requires $\epsilon \gg N^{-1/2}$. So we have[13]

$$I_{gain}(p = 1/2) \approx 0.$$

[13]To be more precise, $I_{gain}(p = 1/2) \sim \log N$, however it's negligible compared to the information gain for $p \neq 1/2$, which turns out to be proportional to N.

We can also consider the case in which $p \approx 0, 1$. That is we can find the volume $\mathcal{V}_\epsilon$ which is bounded between the origin and the surface

$$p_1 + p_2 + \cdots + p_N = N\epsilon.$$

This gives us the *maximum* amount of the information gain. It's easy to see that

$$\mathcal{V}_\epsilon \leq \frac{(\epsilon N)^N}{N!}.$$

So the maximum information gain should be

$$I = \log \mathcal{V}_\epsilon \geq -N \log(\epsilon e)$$

where e is the Neper number. In fact, we can show that $I = -N \log(\epsilon e)$. Let's compare (the absolute value of) of the first and the second term in equation (3.7)

$$t_1/t_0 = N(n-1)^N/n^N = N(1-1/n)^N = N(1-1/(N\epsilon))^N = N \exp(-\frac{1}{\epsilon}).$$

We can choose $\epsilon \sim N^{-1}$. The second term can be then neglected (of course only in this case, i.e. for $p = \epsilon$). This is also true for the rest of the series expansion[14]. So we have

$$I_{max} = -N \log \epsilon e \tag{3.10}$$

For $\epsilon \sim 1/N$ (which is legitimate when $p \approx 0$ or 1), the information gain will be

$$I \approx N \log N$$

[14] We can do the same thing for the m-th term

$$t_m/t_0 < (\frac{eN}{m} e^{-1/\epsilon})^m.$$

Again for $\epsilon \sim N^{-1}$ this would be smaller than $eNe^{-1/\epsilon}$. So the sum of all the terms other than the first divided by the first term would be smaller than

$$N\epsilon . eN \exp(-\frac{1}{\epsilon}) \xrightarrow[N\to\infty]{} 0.$$

Geometrically this means that almost all the contribution to the volume comes from the tiny (super)cube whose side is of the order of ϵ.

up to corrections of the order of N. This is identical to the bound that we found on the maximum information in equation (3.2). Indeed, we see that the bound is actually saturated in this case.

So far we examined the problem in some special cases but were limited by the difficult integral in equation (3.8). In general, we can find the answer in the following way. Assume that we are partitioning the probability space into $M(=\frac{1}{\delta})$ tiny cells. For N particles, there are M^N many ways to distribute the particles into different cells. To find the information gain, we must determine how many different ways there are to distribute the particles subject to

$$\sum_{i=0}^{M-1} n_i p_i \in (Np - N\epsilon, Np + N\epsilon),$$

where n_i is the number of particles in the i-th cell and $p_i = i/M$, as already defined, is the probability of finding these particles in the state $|\uparrow\rangle$. Let's reformulate the problem in the following terms. We would like to find the number of the states constrained to

$$\sum_{i=0}^{M-1} n_i p_i = N\bar{p} \tag{3.11}$$

where $\bar{p} \in (p-\epsilon, p+\epsilon)$. We will later sum over all these $\bar{p}$'s. In the following, we use the method of the maximum entropy [21]. The number of different ways of choosing sets of $n_1, n_2, \ldots, n_M$ particles from N particles is

$$\frac{N!}{n_1! n_2! \ldots n_M!}.$$

The method of maximum entropy teaches us to maximize this expression subject to the constraint (3.11). It is easier to maximize the logarithm of this expression

$$-N \sum_{i=0}^{M-1} x_i \log x_i,$$

where $x_i = n_i/N$. We should then maximize

$$-\sum_{i=0}^{M-1} x_i \log x_i + A\left(\sum_{i=0}^{M-1} x_i p_i - \bar{p}\right) + B\left(\sum_{i=0}^{M-1} x_i - 1\right),$$

where A and B are Lagrange multipliers. By taking the derivative with respect to p_i and equating the result to zero we find

$$x_i = e^{Ap_i + B'}$$

where $B'(= B + 1)$ and A are constant (we drop the prime in the following). Plugging this back in the previous relations, we find

$$-\sum_{i=0}^{M-1} x_i \log x_i = -\sum_{i=0}^{M-1} x_i(Ap_i + B) = -(A\bar{p} + B).$$

The information gain is then

$$\log M^N/\mathcal{N} = N \log M + N(A\bar{p} + B).$$

So the problem reduces to finding the coefficients A and B. Applying the two constraints, we have

$$\sum_{i=0}^{M-1} e^{Ap_i + B} = 1, \quad \sum_{i=0}^{M-1} p_i e^{Ap_i + B} = \bar{p}.$$

Note that $p_i = i\delta$ where $\delta = 1/M$. The first constraint gives us

$$\sum_{i=0}^{M-1} e^{Ap_i + B} = e^B \sum_{i=0}^{M-1} \left(e^{A\delta}\right)^i = e^B \frac{e^A - 1}{e^{A\delta} - 1} = 1$$

The LHS of the second constraint is just the derivative of the LHS of the first constraint with respect to A. Then,

$$e^B \frac{\partial}{\partial A} \frac{e^A - 1}{e^{A\delta} - 1} = \bar{p}.$$

Taking the ratio of the last two relations eliminates B and gives an equation in terms of A

$$\frac{\partial}{\partial A} \log(e^A - 1) - \frac{\partial}{\partial A} \log(e^{A\delta} - 1) = \bar{p}.$$

We are confronted by three possibilities: 1. $A \gg 1$, 2. $A \sim 1$ and 3. $A \ll 1$. It turns out that only the second possibility is self-consistent. With $A\delta \ll 1$, the last relation becomes

$$\frac{1}{1 - e^{-A}} = \frac{1}{A} + \bar{p}. \tag{3.12}$$

This relation is true for the range $(-\infty, +\infty)$ of parameter A. We can solve for B in terms of A

$$B = \log \delta + \log \frac{A}{e^A - 1}.$$

The information gain is then

$$I_{gain} := NK(\bar{p}) = N\left(A\bar{p} + \log(1 + A\bar{p})\right) \tag{3.13}$$

where A should be solved from (3.12) in terms of $\bar{p}$ and plugged in here. We have also defined the function $K = I/N$. Note that in finding $\mathcal{N}$, we didn't sum over all $\bar{p}$ in $(p - \epsilon, p + \epsilon)$ because the number of possibilities grows exponentially in N as $\bar{p}$ gets closer to $1/2$. Therefore, it is safe to plug in $p + \epsilon$ in the last relation when $p < 1/2$ or $p - \epsilon$ when $p > 1/2$. Taking $p < 1/2$, we find

$$I_{gain} = NK(p + \epsilon) = NK(p) + N\epsilon K'(p)$$

where we have Taylor expanded it. The error probability is given by $P_{error} := \gamma \sim 1/N\epsilon^2$. We should choose $\epsilon \ll 1/\sqrt{N\gamma}$ to enforce small error. The second term in the last equation is then much less than $\mathcal{O}\left(\sqrt{N/\gamma}\right)$ and is negligible with respect to the first term (for very large N). So the information gain is

$$I_{gain} = NK(p). \tag{3.14}$$

It is encouraging that the information gain almost does not depend on the error as long as it is much smaller that 1 while it is larger than $1/\sqrt{N}$ (or $1/N$ in the extreme cases). We can then find the information gain purely in terms of $p = n_\uparrow/N$. This information per particle is plotted in Figure (3.2).

Let's examine this expression in a number of limits.

1. p is close to 1 ($1 - p \ll 1$) when $A \to \infty$. In this limit we have $1 = p + \frac{1}{A}$. Then the information gain is

$$I_{gain} = -N \log\left((1 - p)e\right)$$

where e is the Neper number. Note that the uncertainty of p is ϵ and the largest value one can consider for p is $1 - \epsilon$. In this extreme case, this will exactly reproduce the result of equation (3.10). We will also get a similar result when p is close to 0

$$I_{gain} = -N \log\left(pe\right).$$

2. The other extreme is when p is close to the middle of the distribution, i.e. $|p - 1/2| \ll 1$. In this case, we have

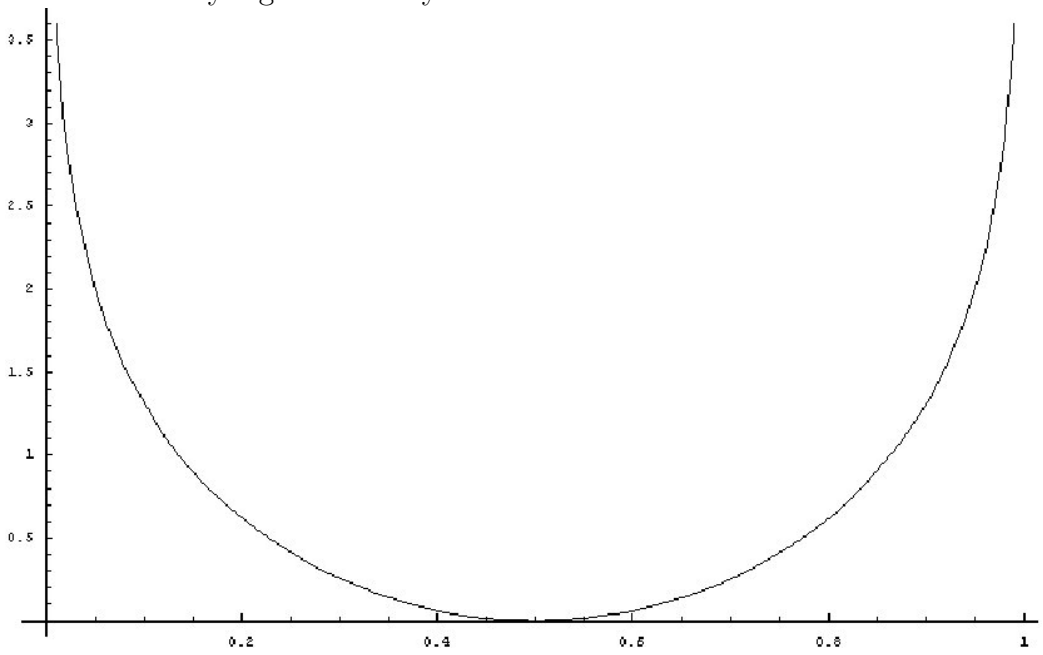
$$I = N \;\; 24(p - 1/2)^2.$$

At the first sight, it seems that there is some discrepancy between this and the result from (3.9). However the latter is only true when $|p-1/2| \sim 1/\sqrt{N}$ while the former is true if the usual corrections of the order of $\log N$ are negligible, which requires that

$$|p - 1/2| \gg \frac{\sqrt{\log N}}{\sqrt{N}},$$

so they are consistent.

Figure 3.2: The information gain divided by N. Note that this function tends to infinity logarithmically at the extremes.



To summarize, we derived the information gain in terms of the mean probability in measuring the spin z of the particles. The best information

we could get from an ensemble of particles in a product state is

$$I = \sup_{\hat{n}} I_{gain}(p_{\hat{n}}). \tag{3.15}$$

So the maximum information will be gained as the maximum $|p_{\hat{n}} - 1/2|$. Suppose that the average density operator is

$$\mathrm{diag}\{\bar{\rho}\} = \begin{pmatrix} p_* & 0 \\ 0 & 1-p_* \end{pmatrix}, \tag{3.16}$$

then the maximum $|p_{\hat{n}} - 1/2|$ is obviously $|p_* - 1/2|$ and the information content is then

$$I = NK(p_*). \tag{3.17}$$

Note that the maximum information is always available to us. Although in the above argument, we assumed that we measure the *individual* spins along a specific axis, we can do better: the average density operator ($\bar{\rho}$) can be determined by a collective measurement as we showed in section 2.2.

It might seem that the maximum information should be larger than (3.17). The argument can be illustrated in Figure (3.3). Knowing "$p = p_*$" geometrically means that the average density operator is somewhere in the dark spot in Figure (3.3a) while knowing (3.16) (i.e. $|\bar{\rho}_{estimate} - \bar{\rho}_{actual}| < \epsilon$) corresponds to the dark spot in Figure (3.3b).

However, we will show that the excess of the information is negligible. The proof again lies in the fact that we should concern ourselves only with the region which is closer to the center of the Bloch sphere. That is the number of states decreases as we move away from the center of the Bloch sphere. Suppose that the number of states in the small region of size ϵ^2 in Figure (3.3b) is $\mathcal{V}(p)$. Then the number of states in Figure (3.3a) is definitely less than $\mathcal{V}(p).1/\epsilon^2$. So the difference of the information is

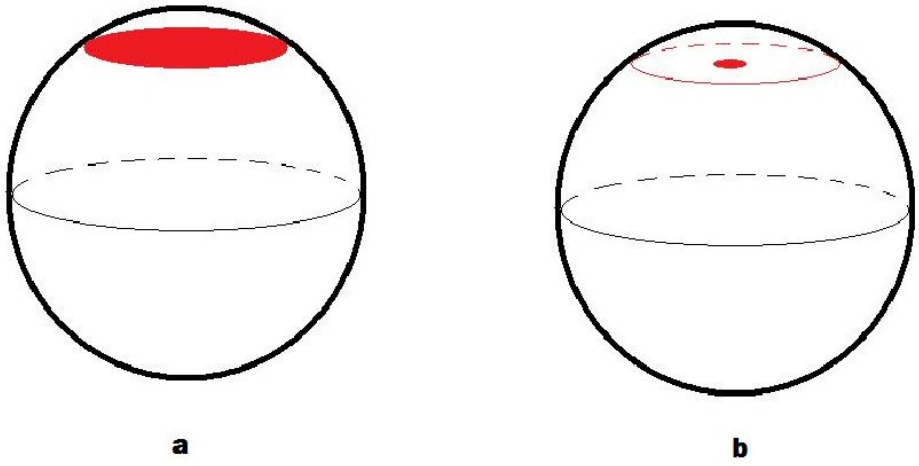
$$\log \frac{1}{\mathcal{V}(p)} - NK(p) < \log(1/\epsilon^2).$$

Since $\epsilon \sim 1/\sqrt{N}$ at worst, the difference turns out to be logarithmic in N.

So perhaps the most important result of this letter can be stated as

$$I_{gain} = NK(p_*) \;\; \text{where } p_* = \frac{1}{2} + \frac{1}{2}\sqrt{2\mathrm{Tr}(\bar{\rho}^2) - 1} \tag{3.18}$$

Figure 3.3: a) The information corresponding to $p = p_{\hat{n}}$. b) The total information.



for a two-level system. Note that the range of p is $(1/N, 1-1/N)$ because p is only found with some resolution which can not be less than $1/N$.

This equation is not very similar to the well-known definitions of the information gain. For example, the information gain per particle (for a two-level system) could be

$$I(\bar{\rho})/N \gg 1.$$

Even in the extreme case

$$I_{max}/N = \log N.$$

However the fact that the information gain per particle is much more than 1 does not contradict the quantum information theory in any way. In the context of information theory, the information is mostly concerned with the information coding, specifically the number of different letters we can use in order to encode some message. Here we can easily calculate the number of states that can be discriminated. We can find the average density operator with some resolution $|\bar{\rho}_{estimate} - \bar{\rho}_{actual}| < 1/\sqrt{N}$. So we can discriminate

$$(\frac{1}{1/\sqrt{N}})^{\alpha'} = N^{\alpha},$$

different states ($\alpha \sim 1$). The logarithm of this number is

$$\alpha \log N \quad (< N \text{ bits})$$

which is (much) less than N. Therefore there is no contradiction.

In the next chapter, we summarize the results and investigate the implications.

Chapter 4

Conclusion and Discussion

The information gain that we found here is different from Shannon information or its close cousin von Neumann information (entropy). In this chapter we argue more extensively why we formulated a different notion of the information gain and didn't apply the usual definitions of the information from Quantum Information theory(QI).

First of all, the problem that we studied in this letter involved infinitely many states (all states of a Bloch sphere, for example). In the context of Quantum Information theory, however, we usually consider finite number of states (an ensemble of states $\{\rho_x, p_x\}$ where x is usually assumed to take finite different values). In Appendix A, we argue that the Shannon Information would not be straightforward for infinite number of states.

In the following, we will give a brief summary that how the accessible information is defined in the context of the Quantum Information theory[22] and consider the preassumptions more carefully. Assume that we have an ensemble of states $\mathcal{E} = \{\rho_x, p_x\}$ where p_x is the *a priori* probability that the state ρ_x is chosen or sent on a communication channel. The receiver can then collect some information by performing a generalized measurement, the POVM $\{M_y\}$. If some state ρ_x is sent, he will find the outcome y with the following conditional probability

$$p(y|x) = \mathrm{Tr}(M_y \rho_x).$$

The conditional probabilities determine the amount of information that he can gain on the average, the mutual information $I(X:Y)$ of the preparation and the measurement outcome. The maximum information gain is then

$$Acc(\mathcal{E}) = \max_{\{M_y\}} I(X:Y).$$

It might seem that we studied a similar problem except that we considered any (possibly mixed) state in the Hilbert space, i.e. all $\rho_x \in \mathcal{H}_{one-particle}$ and the *a priori* probabilities were assumed to be all equal. Discretizing the

Hilbert space, we then have $\mathcal{E} = \{\rho_x, p_x = 1/N\}$ where N is the number of the states in the discretized Hilbert space.

This line of thought has been pursued, for example, by Asher Peres. In his book [13], he gives the following example

"Suppose that the only information prior to a test of σ_z is that the initial state was pure. It satisfies $\vec{\sigma}.\mathbf{n}|\psi\rangle = |\psi\rangle$ with equal probabilities for all directions of the unit vector $\mathbf{n}$..."

Peres discretizes the Hilbert space to N small intervals and assumes the *a priori* probabilities to be $p = 1/N$. He defines the *posteriori* information by Bayes's theorem in accordance with the result of the measurement. The information gain he finds is then $I_{ave} = 0.19$.

In our viewpoint, this approach is not appropriate. There is a major difference between the Quantum-Information theoretic approach and ours. In the former, the probabilities p_x's are the *actual* probabilities. That is, a fraction p_x of the states (that are drawn from the ensemble) are *really* in the state ρ_x. This allows us to consider only a subspace (the typical subspace) of the whole Hilbert space while discarding the rest of it. Specifically this means that for a large sample, the average of the density operators of the states ($\sum \rho_i$ divided by the total number of them) *should* be $\sum p_x \rho_x$. The problem that we studied in this paper is totally different. Here we assumed that all states are equally probable because we did not know any better. $p_x = 1/N$ doesn't mean that the states are *truly* distributed with such probability or the average density operator is $\frac{1}{d}\mathbf{1}$; a total section was actually devoted to find the average density operator. In QI, one assumes the probabilities *a priori* and then tries to find the information conditioned to these probabilities. In our problem we don't know the *a priori* probabilities and, as we showed, we can only find the average of the density matrix while it's already taken for granted in the context of QI.

As we have already pointed out there is no way to get any information about the quantum state unless we've got some *a priori* knowledge of the system. In this letter we assumed that the state of the system is a product state of pure states. It would be very interesting to ask the same question given different *a priori* knowledge.

It's very curious to observe that the information is gained only if we know something in advance. As discussed in the first chapter, we might detect a new star far away and gain some information about it. As we outlined in

the first chapter, one explanation can be the continuous measurement. That is, we may assume that the star's position is continuously measured by its environment. What the measurement does is simply to project down the system onto some proper subspace of the Hilbert space. In this example, this is the position basis. The question is then whether we always need some sort of measurement to provide us with some *a priori* knowledge. Can the unitary evolution (as opposed to the measurement) do this task? We argue that it cannot.

Suppose we could evolve any state (of the Hilbert space) into a strict subspace of the Hilbert space. Defining an invariant measure of the Hilbert space, the volume of this subspace is $\mathcal{V}' = \int_{\Omega'} D[\psi]$. Since the measure of the volume is the same under unitary evolution, the volume must remain the same. So $\mathcal{V} = \mathcal{V}'$, where $\mathcal{V}$ is the total volume of the Hilbert space. That is we can't project the whole Hilbert space onto any strict subspace of it. This is rather trivial and, in fact, is similar to the argument that the volume of the phase-space remains the same in classical mechanics (that is $\int dpdqf(p,q) = const$).

So it seems that (non-unitary) measurement-like processes are essential in order to gain any sort of information in quantum mechanics.

In this letter we showed that even for a system of many particles in some unknown (but product) state, we can gain much information and we hope this could partially answer the question that how the information gain is possible from the unknown environment.

Bibliography

[1] M. A. Nielsen and I. L. Chuang, Quantum computation and quantum information, Cambridge University Press, 2000.

[2] C. Brukner and A. Zeilinger, Phys. Rev. Lett. 83, 3354 (1999).

[3] H. Everett, Rev. Mod. Phys. 29, 454-462 (1957).

[4] B. S. M. De Witt, Quantum mechanics and Reality, Physics Today 23, No. 9, pp. 30-35 (1970).

[5] C. M. Caves, C. A. Fuchs, R. Schack, Subjective probability and quantum certainty (2006), arXiv:quant-ph/0608190.

[6] Y. Aharonov, J. Anandan, L. Vaidman, Phys. Rev. A 47 4616(1993).

[7] W. G. Unruh, Reality and measurement of the wave function, Phys. Rev. A 50, 882 (1994).

[8] O. Alter and Y. Yamamoto, Quantum Measurement of a Single System, Wiley, New York, 2001.

[9] U. Leonhardt, Measuring the Quantum State of Light, Cambridge University Press, Cambridge, England, 1997.

[10] S.P. Walborn et al., Nature 440, 1022(2006).

[11] S. J. van Enk, e-print arXiv:quant-ph/0606017.

[12] T. Bhattacharya, S. Habib, and K. Jacobs, Phys. Rev. Lett. 85, 4852 (2000).

[13] A. Peres, Quantum Theory: Concepts and Methods, Kluwer Academic, Dordrecht, 1993.

[14] J. Hartle, Quantum mechanics of individual systems. Am. J. Phys., 36, 704712 (1968).

[15] E. Farhi, J. Goldstone, S. Gutmann, How probability arises in quantum-mechanics. Annal. phys., 192, 368382 (1989).

[16] D. Duetsch, Proc. R. Soc. London A 455 (1999) 3129.

[17] D. Wallace, Quantum probability from subjective likelihood: Improving on Deutschs proof of the probability rule, Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys, 38, 311-332 (2007).

[18] S. Massar and S. Popescu, Phys. Rev. Lett. 74, 1259 (1995)

[19] J. Kofler, C. Brukner, Phys. Rev. Lett. 99, 180403 (2007)

[20] G. Polya, Berechnung eines Bestimmten Integrals, Math. Ann. 74 (1913) 204-212.

[21] E. T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.

[22] J. Preskill, Lecture notes for physics 229: Quantum information and computation, URL http://www.iqi.caltech.edu.

Appendix A

Shannon Information

In the following, we derive Shannon information in the usual case where there are only finite number of outcomes and argue the generalization to the infinitely many states is not straightforward. We will then show Shannon information is a special case of the information gain as defined in this letter.

Consider a random variable which might assume some value from a set of L different possibilities. It might assume the first one with probability p_1, the second one with probability p_2, etc. For large N number of events, $p_1 N$ of the outcomes will assume the first value, $p_2 N$ assume the second value, etc. (typical case). There is a small probability that this might not be the case that is called the *atypical* case. We can ignore the atypical states for large enough N. There are

$$\mathcal{N}_{typical} = \frac{N!}{(Np_1)!(Np_2)!\dots(Np_L)!}.$$

typical states that may occur with equal probability. The uncertainty of the outcomes is then

$$\log \mathcal{N}_{typical} \approx NH(p_1, p_2, \dots, p_L)$$

where H is the Shannon entropy defined as

$$H(p_1, p_2, \dots, p_L) = -\sum p_i \log p_i.$$

It might seem that this procedure is also applicable to the continuous case in which there are infinitely many states ($L = \infty$). However the *typical subspace* arises for N large enough so that the mean number of events per state will be much larger than 1, i.e. $N/L \gg 1$. For $L = \infty$, this would be never satisfied.

In the following, we show that the definition of the information gain, as presented in this letter, would result in the Shannon information for an ensemble $\{x, p_x\}$ where x is the (classical state) and p_x is the corresponding *a priori* probability. The "likely" space, as defined in this letter, consists of

the "typical sequences" in this case. Before the "measurement" the number $\mathcal{N}$ of typical sequences is

$$2^{n(H+\delta)} \geq \mathcal{N}(\epsilon, \delta) \geq (1-\epsilon)2^{n(H-\delta)}$$

where $P_{error} = \epsilon$ is the probability of error. In our language, this number is the same as $\mathcal{N}_{total}$. That is, we have

$$\mathcal{N}_{total} = \mathcal{N}(\epsilon, \delta).$$

After performing the measurement (reading the sequence of the letters), we find one sequence. Therefore $\mathcal{N}_{measured} = 1$. Note that ϵ and δ can be assumed as small as desired for large enough N. Following our definition of the information gain, we then find

$$I_{gain} = NH(x).$$

where the error probability tends to zero for $N \to \infty$.

Appendix B

Law of the Large Numbers

Assume that $X_1, X_2, \ldots, X_N$ are N random variables having possibly different distributions. In the following, we prove an extension of the law of large numbers[15]. The argument is almost along the same lines as the standard law.

Define the operator E which takes the average of a random variable

$$\begin{aligned} E(S_N^2) &= \sum_{i,j=1}^{N} \frac{E(X_i X_j)}{N^2} \\ &= \frac{1}{N^2} \sum_{i \neq j} E(X_i X_j) + \frac{1}{N^2} \sum_i E(X_i^2). \end{aligned}$$

The first term could be simplified as

$$\sum_{i \neq j} E(X_i X_j) = \sum_j (\sum_{i \neq j} E(X_i)) E(X_j)$$

$$= \sum_j (N \bar{E}_N - E(X_j)) E(X_j) = N^2 \bar{E}_N^2 - \sum_{j=1}^{N} E(X_j)^2 = N^2 \bar{E}_N^2 - N \bar{E^2}_N$$

where we defined

$$\bar{E}_N = \frac{1}{N} \sum_{i=1}^{N} E(X_i), \quad \bar{E^2}_N = \frac{1}{N} \sum_{i=1}^{N} E(X_i)^2$$

Plugging this back in the first equation, we find

$$E(S_N^2) = \bar{E}_N^2 + \frac{1}{N^2} \sum_{j=1}^{N} (E(X_j^2) - E(X_j)^2)$$

[15] The law of large numbers applies to N identically distributed random variables.

If the random variable may assume only finite possible outcomes, then the quantity in the bracket is bound from above. Defining

$$\sigma^2 := \frac{1}{N}\sum E(X_j^2) - E(X_j)^2,$$

we have

$$E(S_N^2) = \bar{E}_N^2 + \frac{1}{N}\sigma^2.$$

We can rewrite this as

$$E\left((S_N - E_N)^2\right) = \frac{1}{N}\sigma^2$$

This can be also defined in terms of the probability measure

$$E\left((S_N - E_N)^2\right) = \int dP \;\; (S_N - E_N)^2$$

where dP is the probability measure

$$\int dP \;\; (S_N - E_N)^2 \geq \epsilon^2 p(|S_N - E_N| > \epsilon).$$

$p()$ stands for the probability. So we have

$$p(|S_N - E_N| > \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}. \tag{B.1}$$

Therefore, for $N \to \infty$, we find $S_N \to E_N$.

We can make a stronger result as follows. For the sake of simplicity, suppose there are only two values that X may assume: 0, 1. We have

$$\int dP(X - E(X)) = 0 \Rightarrow \int_{X=0}^{E(X)} dP(E(X) - X) = \int_{X=E(X)}^{1} dP(X - E(X)).$$

We then find

$$\int dP(X - E(X))^2 \leq \int dP|X - E(X)|$$
$$= 2\int_{X=0}^{E(X)} dP(E(X) - X) = 2\int_{X=E(X)}^{1} dP(E(X) - X) \leq \min\left(E(X), 1 - E(X)\right).$$

Since $E(X)$ is only determined up to some uncertainty ϵ, we have

$$\int dP(X - E(X))^2 \leq \min\left(S(X), 1 - S(X)\right) + \epsilon.$$

This would be specially useful when the average is very close to one extreme of $[0, 1]$. Putting all this together, we find

$$p(|S_N - E_N| > \epsilon) \leq \frac{\bar{E_N^2}}{N\epsilon^2} \leq \frac{\min\left(S(X), 1 - S(X)\right)}{N\epsilon^2} + \frac{1}{N\epsilon} \tag{B.2}$$

In the extreme case that all the outcomes (but possibly a few of them) turn out to be 0 or 1, we find

$$P_{error} \sim \frac{1}{N\epsilon}.$$

For the probability to be reasonably small, we must have

$$\epsilon \sim 1/N.$$

Appendix C

Central Limit Theorem

Let X_1, X_2, X_3, ... be a set of n independent and identically distributed random variables having finite mean values and variance $\sigma^2 > 0$. The central limit theorem (also known as the second fundamental theorem of probability) states that as the sample size n increases, the distribution of the sample average approaches the normal distribution with a mean μ and variance σ^2/n irrespective of the shape of the original distribution.

Let the sum of the random variables be S_n, given by $S_n = X_1 + ... + X_n$. Then, defining

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}\,,$$

the distribution of Z_n converges towards the standard normal distribution $N(0,1)$ as n approaches ∞. This can be also reformulated as

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}},$$

where

$$\overline{X}_n = S_n/n = (X_1 + \cdots + X_n)/n$$

is average of the outcomes

Appendix D

Lyapunov Condition

Let X_n, $n \in \mathbf{N}$, be a sequence of independent random variables. Suppose that each X_n has finite expected value $\mathbf{E}[X_n] = \mu_n$ and finite variance $\mathrm{Var}[X_n] = \sigma_n^2$. Suppose also that the third central moments

$$r_n^3 := \mathbf{E}[\left|X_n - \mu_n\right|^3]$$

are finite and satisfy the Lyapunov condition

$$\lim_{N\to\infty} \frac{\left(\sum_{n=1}^N r_n^3\right)^{1/3}}{\left(\sum_{n=1}^N \sigma_n^2\right)^{1/2}} = 0.$$

Let the random variable $S_N := X_1 + \cdots + X_N$ denote the N-th partial sum of the random variables X_n. Then the normalized partial sum

$$Z_N := \frac{S_N - \sum_{n=1}^N \mu_n}{\left(\sum_{n=1}^N \sigma_n^2\right)^{1/2}}$$

converges in distribution to a standard normal random variable as $N \to \infty$. Less formally, for "large" N, S_N is approximately normally distributed with expected value

$$\mathbf{E}[S_N] \approx \sum_{n=1}^N \mathbf{E}[X_n]$$

and variance

$$\mathrm{Var}[S_N] \approx \sum_{n=1}^N \mathrm{Var}[X_n].$$