

Quantum non demolition measurements: Parameter estimation for mixtures of multinomials

Tristan Benoist, Fabrice Gamboa and Clément Pellegrini

*Institut de Mathématiques de Toulouse UMR5219, Université de Toulouse ; CNRS,
Équipe de Statistique et Probabilités,
Université Paul Sabatier, 31062 Toulouse Cedex 9, France*
e-mail: tristan.benoist@math.univ-toulouse.fr
e-mail: fabrice.gamboa@math.univ-toulouse.fr
e-mail: clement.pellegrini@math.univ-toulouse.fr

Abstract: In Quantum Non Demolition measurements, the sequence of observations is distributed as a mixture of multinomial random variables. Parameters of the dynamics are naturally encoded into this family of distributions. We show the local asymptotic mixed normality of the underlying statistical model and the consistency of the maximum likelihood estimator. Furthermore, we prove the asymptotic optimality of this estimator as it saturates the usual Cramér Rao bound.

Received July 2017.

Contents

1	Introduction	555
2	Mixture of multinomials	557
3	Local asymptotic mixed normality	560
4	Maximum likelihood estimation	562
	4.1 Consistency	562
	4.2 Saturation of Cramér–Rao bound	563
5	Applications to quantum non-demolition measurement	565
A	Uniform convergence of logarithm of sums of uniformly convergent sequences	569
	Acknowledgments	570
	References	570

1. Introduction

Measuring directly a small *quantum sized* physical system is done by letting it interact with a macroscopic instrument. This procedure can result in the destruction of the measured system. For example photons are absorbed to create an electronic signal. To avoid the destruction of the measured system, one relies then on indirect measurements. The system first interacts with an auxiliary system, or probe, that is then measured. The goal is then to infer the

system state from the information obtained through this indirect measurement. Though, from the laws of quantum mechanics, this procedure induces a back action on the system that may change its state. Moreover, the measurement outcome being inherently random, the system state may become random itself. Hence, if one aims at measuring indirectly a physical quantity of the system, the indirect measurement must be tuned such that, if the system state corresponds to an almost sure value of the physical quantity of interest, then, indirectly measuring it will not modify its state. This kind of indirect measurements is called Quantum Non Demolition (QND) measurements. It has been introduced in the eighties as a technique for precise measurements [6]. Maybe one of the experiment illustrating best QND measurements is Haroche's group one. By sending atoms through a supra conducting cavity containing a monochromatic electromagnetic field, and measuring the atoms, it is possible to measure the number of photons inside the cavity without destroying them [8].

To increase the amount of information on the system state obtained through a QND measurement, the procedure is repeated. The system evolution is then described by an unobserved Markov chain (ϕ_n) (see Section 5 for the complete description). The only handy observation is the sequence of the measurement results denoted by (X_n) . Baye's law maps the information of (X_n) in the evolution of the Markov chain.

In general, the sequence of random variables (X_n) is not i.i.d. (even not Markovian). Therefore, statistical inference for QND measurement cannot fully rely on standard results on i.i.d. models. Efficient parameter estimation is of course crucial for these experiments. Particularly if one hopes to have a faithful estimation of the system state Markov chain (ϕ_n) .

In this paper, we show that the QND model enters perfectly in the framework of the usual statistical asymptotic theory. More precisely, we provide a complete study in terms of local asymptotic mixed normality (LAMN) of the model (we refer to [14] for the whole theory). It is worth noticing that a quantum analog of local asymptotic normality (LAN) called quantum local asymptotic normality (QLAN) has been developed in the context of quantum statistics [9, 12]. Similarly quantum extensions of classical notions such as Quantum Fisher Information and Cramér Rao bound have been developed [11, Section 2.2.5]. Here, we will not follow this approach and will concentrate on more classical statistical properties. Our results rely on the fact that our model, thanks to the QND condition, is actually a mixture of i.i.d. statistical models. More precisely, it has been show in [2, 3, 4, 5, 1] that the probability space describing these experiments can be divided into asymptotic events (belonging to the tail algebra) such that (X_n) conditioned to one of such asymptotic event is a sequence of i.i.d. random variables. So that, the law of (X_n) is a mixture of i.i.d. laws. The weights involved into the mixture depends on the initial state of the system.

The conditioning making (X_n) an i.i.d. sequence is highly exploited in order to derive the LAMN property. To our knowledge this is the first time that the LAMN property is shown in this context. After proving the LAMN property we study the maximum likelihood estimation and prove that it is optimal in the sense that the Cramér Rao bound is achieved asymptotically. Note that para-

metric estimation for indirect measurements has been previously investigated in [10, 7, 13] with different assumptions. But, the theory of asymptotic likelihood has not been studied therein.

The paper is organised as follows. In Section 2, we discuss the model of multinomial mixture studied along the paper. In Section 3 we show the local asymptotic mixed normality. Section 4 is devoted to the results for the maximum likelihood estimator (consistency and saturation of the Cramér Rao bound). Finally in Section 5 we work on the QND model underlining the link with multinomial mixtures. Further, some numerical simulations illustrate our results on a QND toy model inspired by [8].

2. Mixture of multinomials

Let $\mathcal{A} = \{1, \dots, l\}$, $\mathcal{P} = \{1, \dots, d\}$ and Θ be a compact subset of \mathbb{R}^D with a non empty interior. For any $\alpha \in \mathcal{P}$ and $\theta \in \Theta$, the quantity $(p_\theta(j|\alpha))_{j \in \mathcal{A}}$ denotes a probability distribution over \mathcal{A} , that is $p_\theta(j|\alpha) \in (0, 1)$ and $\sum_{j \in \mathcal{A}} p_\theta(j|\alpha) = 1$. In the sequel for any $j \in \mathcal{A}$ and $\alpha \in \mathcal{P}$ the notation $p_\cdot(j|\alpha)$ holds for the function $\theta \mapsto p_\theta(j|\alpha)$.

Let \mathbb{P} be a probability distribution, we will use the notation $\stackrel{\mathcal{L}-\mathbb{P}}{=}$ to mean equality in distribution when the underlying probability space is endowed with \mathbb{P} . This notation will also be used for convergence in distribution writing sometimes $\stackrel{\mathcal{L}-\mathbb{P}_n}{=}$ when a family of probability measures (\mathbb{P}_n) is involved. In the paper the notation $\mathcal{N}(m, \sigma^2)$ is used for the Gaussian distribution of mean m and variance σ^2 . For any $x \in \mathbb{R}^D$ and any subset $A \subset \mathbb{R}^D$, the set $A + x$ will denote $A + x = \{y + x, y \in A\}$.

Let us now describe the probability model that we will study. Let $\Omega = \mathcal{A}^{\mathbb{N}}$ and let \mathcal{F} be the smallest σ -algebra containing the cylinder sets $\{\omega \in \Omega | \omega_k = j_k, \forall k \leq n\}$. All the measures introduced afterwards are defined on the measurable space (Ω, \mathcal{F}) without mentioning it.

For any $\theta \in \Theta$ and for each $\alpha \in \mathcal{P}$, let $\mathbb{P}_{\theta|\alpha}$ be the multinomial probability measure built on the weights $(p_\theta(j|\alpha))_{j \in \mathcal{A}}$. Namely, for any n -tuple $(j_1, \dots, j_n) \in \mathcal{A}^n$,

$$\mathbb{P}_{\theta|\alpha}(j_1, j_2, \dots, j_n) = \prod_{k=1}^n p_\theta(j_k|\alpha).$$

Let $(q(\alpha))_{\alpha \in \mathcal{P}}$ be a probability measure on \mathcal{P} , we denote by \mathbb{P}_θ^q the probability measure defined as a convex combination of the measure $\mathbb{P}_{\theta|\alpha}$ with weights $(q(\alpha))_{\alpha \in \mathcal{P}}$:

$$\mathbb{P}_\theta^q = \sum_{\alpha \in \mathcal{P}} q(\alpha) \mathbb{P}_{\theta|\alpha}.$$

Without loss of generality, we shall always assume that $q(\alpha) > 0$ for all $\alpha \in \mathcal{P}$. Indeed, one can reduce the set \mathcal{P} if needed.

In the sequel we shall use the process $(N_n(j))_{j \in \mathcal{A}}$ where for all $n \in \mathbb{N}$ and all $j \in \mathcal{A}$

$$N_n(j)(\omega) = \sum_{k=1}^n \mathbf{1}_{\omega_k=j},$$

for all $\omega \in \Omega$, which counts the number of times the result j appears before time n . Remark that $\sum_{j \in \mathcal{A}} N_n(j) = n$. The strong law of large number for i.i.d L^1 random variables implies that

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = p_\theta(j|\alpha), \quad \mathbb{P}_{\theta|\alpha} - a.s., \tag{1}$$

for all $\alpha \in \mathcal{P}$, all $j \in \mathcal{A}$ and all $\theta \in \Theta$.

We study the statistical model $(\mathbb{P}_\theta^q, \theta \in \Theta)$. As we shall see, under the identifiability condition below, our results will be independent of q . Hence, in the sequel we shall alleviate the notation replacing \mathbb{P}_θ^q by \mathbb{P}_θ .

Our main assumption on the different multinomials is the following:

Assumption ID: For any (α, θ) and $(\beta, \theta') \in \mathcal{P} \times \Theta$ such that $(\alpha, \theta) \neq (\beta, \theta')$, $\exists j \in \mathcal{A}$ such that,

$$p_\theta(j|\alpha) \neq p_{\theta'}(j|\beta).$$

Remark: If for every $\theta \in \Theta$, Assumption **ID** does not hold for two couples (α, θ) and (β, θ) , it can be enforced by reducing \mathcal{P} , identifying the elements giving the same distributions $p_\theta(\cdot|\alpha/\beta)$.

From now on, we will denote by $\theta^* \in \Theta$ the true value of the parameter θ . We assume that θ^* is in the interior of Θ . The following definition will be useful.

Definition 1. For any $\theta^* \in \Theta$ and $\gamma \in \mathcal{P}$, let,

$$\Omega_{\theta^*|\gamma} := \text{supp } \mathbb{P}_{\theta^*|\gamma} \quad \text{and} \quad \Gamma_{\theta^*} := \sum_{\gamma \in \mathcal{P}} \gamma \mathbf{1}_{\Omega_{\theta^*|\gamma}}.$$

When it does not lead to confusion we may omit the index θ^* for both the sets $\Omega_{\theta^*|\gamma}$ and the random variable Γ_{θ^*} .

Proposition 1. Assume that **ID** holds. Then, for any $\theta^* \in \Theta$,

$$\mathbb{P}_{\theta^*}(\Omega_{\theta^*|\gamma}) = \mathbb{P}_{\theta^*}(\Gamma_{\theta^*} = \gamma) = q(\gamma).$$

Proof. Assumption **ID** implies that for any two distinct $\alpha, \beta \in \mathcal{P}$, there exists $j \in \mathcal{A}$ s.t. $p_{\theta^*}(j|\alpha) \neq p_{\theta^*}(j|\beta)$. Eq. (1) implies that for any $\alpha \in \mathcal{P}$,

$$\Omega_{\theta^*|\alpha} \subset \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} N_n(j) = p_{\theta^*}(j|\alpha), \forall j \in \mathcal{A} \right\}.$$

It follows that $\Omega_{\theta^*|\alpha} \cap \Omega_{\theta^*|\beta} = \emptyset$ if $\alpha \neq \beta$. Hence,

$$\mathbb{P}_{\theta^*|\alpha}(\Omega_{\theta^*|\beta}) = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

The proposition follows then from the definitions of Γ_{θ^*} and \mathbb{P}_{θ^*} . □

At this stage we need to introduce some quantities quantifying the information and proximity in our models. In particular, we shall use many times the Shannon entropy given a parameter θ and the Kullback–Leibler divergence given θ with respect to θ' . For $\alpha, \beta \in \mathcal{P}$ and $\theta, \theta' \in \Theta$, let

$$S_\theta(\alpha) := - \sum_{j \in \mathcal{A}} p_\theta(j|\alpha) \ln p_\theta(j|\alpha), \tag{2}$$

be the Shannon entropy and

$$S_{\theta|\theta'}(\alpha|\beta) := \sum_{j \in \mathcal{A}} p_\theta(j|\alpha) (\ln p_\theta(j|\alpha) - \ln p_{\theta'}(j|\beta)), \tag{3}$$

be the Kullback–Leibler divergence. In Eq (3), when $\theta = \theta'$ we just write S_θ , that is for $\alpha, \gamma \in \mathcal{P}$ and $\theta \in \Theta$

$$S_\theta(\alpha|\gamma) = S_{\theta|\theta}(\alpha|\gamma) = \sum_{j \in \mathcal{A}} p_\theta(j|\alpha) \ln \frac{p_\theta(j|\alpha)}{p_\theta(j|\gamma)}. \tag{4}$$

Remark: Note that assumption **ID** ensures that (3) and (4) are strictly positive.

We now state a technical lemma which is a key tool to all our proofs (the arguments are closely related to the one used in [3]).

Lemma 1. *Assume that **ID** holds.*

(a) **Almost sure convergence** *Let (X_n^γ) , be a sequence of random variables depending on $\gamma \in \mathcal{P}$. If for any $\gamma \in \mathcal{P}$*

$$\lim_n X_n^\gamma = X^\gamma, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Then

$$\lim_n X_n^{\Gamma_{\theta^*}} = X^{\Gamma_{\theta^*}}, \quad \mathbb{P}_\theta - a.s$$

(b) **Convergence in distribution** *Let (θ_n^*) be a sequence in Θ and (X_n^γ) as in (1). If for any $\gamma \in \mathcal{P}$*

$$\lim_n X_n^\gamma \stackrel{\mathcal{L}-\mathbb{P}_{\theta_n^*|\gamma}}{=} X^\gamma,$$

then,

$$\lim_n X_n^{\Gamma_{\theta_n^*}} \stackrel{\mathcal{L}-\mathbb{P}_{\theta_n^*}}{=} X^\Gamma,$$

where Γ is a r.v. whose distribution is given by $\Pr(\Gamma = \gamma) = q(\gamma)$.

Proof. Assume that (X_n^γ) converges almost surely towards X^γ w.r.t $\mathbb{P}_{\theta^*|\gamma}$. Then,

$$\mathbb{P}_{\theta^*|\gamma}(\cap_N \cup_{n_0} \cap_{n \geq n_0} \{|X_n^\gamma - X^\gamma| < 1/N\}) = 1.$$

Since $\Gamma = \gamma$ $\mathbb{P}_{\theta^*|\gamma}$ -a.s.,

$$\mathbb{P}_{\theta^*|\gamma}(\cap_N \cup_{n_0} \cap_{n \geq n_0} \{|X_n^\Gamma - X^\Gamma| < 1/N\}) = 1.$$

This is true for any $\gamma \in \mathcal{P}$. Since \mathbb{P}_{θ^*} is convex combination of the measures $\mathbb{P}_{\theta^*|\gamma}$,

$$\mathbb{P}_{\theta^*}(\cap_N \cup_{n_0} \cap_{n \geq n_0} \{|X_n^\Gamma - X^\Gamma| < 1/N\}) = 1$$

and (1) holds.

Assume that (X_n^γ) converges weakly towards X^γ w.r.t. $(\mathbb{P}_{\theta_n^*|\gamma})$. Then,

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\theta_n^*|\gamma}(f(X_n^\gamma)) = \mathbb{E}(f(X^\gamma))$$

for any continuous and bounded function f . Since $\mathbb{P}_{\theta_n^*|\gamma}(\Gamma_{\theta_n^*} = \gamma) = 1$ and $\mathbb{P}_{\theta_n^*} = \sum_{\alpha \in \mathcal{P}} q(\alpha) \mathbb{P}_{\theta_n^*|\alpha}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\theta_n^*}(f(X_n^{\Gamma_{\theta_n^*}})) = \sum_{\alpha \in \mathcal{P}} q(\alpha) \mathbb{E}(f(X^\alpha)).$$

That convergence yields (2). □

3. Local asymptotic mixed normality

We first prove that the statistical model $(\mathbb{P}_\theta, \theta \in \Theta)$ is asymptotically equivalent to a mixture of Gaussian models. Let us first recall the definition of local mixed asymptotic normality that we shall prove for our model.

Definition 2 (Local Asymptotic Mixed Normality (LAMN)[14]). *A sequence of statistical models $(\mathbb{Q}_\theta|_{\mathcal{F}_n}|\theta \in \Theta)$ is said to be LAMN at θ^* if, there exists Δ_{θ^*} and J_{θ^*} two random variables such that for any $h \in \Theta - \theta^*$*

$$\lim_{n \rightarrow \infty} \ln \frac{\mathbb{Q}_{\theta^*+h/\sqrt{n}}(\omega_1, \dots, \omega_n)}{\mathbb{Q}_{\theta^*}(\omega_1, \dots, \omega_n)} \stackrel{\mathcal{L}-\mathbb{Q}_{\theta^*}}{=} h^\top \Delta_{\theta^*} - \frac{1}{2} h^\top J_{\theta^*} h,$$

where the law of Δ_{θ^*} conditioned on $J_{\theta^*} = J$ is $\mathcal{N}(0, J)$.

The LAMN property for our model follows from next lemma. It reduces the problem to local asymptotic normality for i.i.d. multinomials.

Lemma 2. *Assume that ID holds and that for any $(j, \alpha) \in \mathcal{A} \times \mathcal{P}$ the function $p(j|\alpha)$ is continuous in θ^* . Let $(\theta_n) \in \Theta$ be a sequence of random variables such that $\lim_n \theta_n = \theta^*$, $(\mathbb{P}_{\theta_n^*|\gamma} - a.s)$. Then,*

$$\left| \frac{\mathbb{P}_{\theta_n}(\omega_1, \dots, \omega_n)}{q(\gamma) \mathbb{P}_{\theta_n^*|\gamma}(\omega_1, \dots, \omega_n)} - 1 \right| = o(n^{-1/2}), \quad \mathbb{P}_{\theta_n^*|\gamma} - a.s. \tag{5}$$

Proof. We prove the stronger result that the convergence is exponential but we will only need the convergence at order $o(n^{-1/2})$.

From the definition of \mathbb{P}_{θ_n} ,

$$\frac{\mathbb{P}_{\theta_n}(\omega_1, \dots, \omega_n)}{\prod_{k=1}^n p_{\theta_n}(\omega_k|\gamma)} = q(\gamma) + \sum_{\alpha \neq \gamma} q(\alpha) \exp \left(\sum_{j \in \mathcal{A}} N_n(j)(\omega) \ln \frac{p_{\theta_n}(j|\alpha)}{p_{\theta_n}(j|\gamma)} \right).$$

The strong law of large numbers (1) and the continuity assumption on the functions $p_{\cdot}(j|\alpha)$ imply

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j \in \mathcal{A}} N_n(j) \ln \frac{p_{\theta_n}(j|\alpha)}{p_{\theta_n}(j|\gamma)} = -S_{\theta^*}(\gamma|\alpha), \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Recall that Assumption **ID** implies $S_{\theta^*}(\gamma|\alpha) > 0$ for any $\alpha \neq \gamma$. Therefore, there exists $s > 0$ independent of γ, α and of the sequence (θ_n) such that,

$$\lim_{n \rightarrow \infty} e^{ns} \sum_{\alpha \neq \gamma} q(\alpha) \exp \left(\sum_{j \in \mathcal{A}} N_n(j) \ln \frac{p_{\theta_n}(j|\alpha)}{p_{\theta_n}(j|\gamma)} \right) = 0, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Now,

$$\frac{\mathbb{P}_{\theta_n}(\omega_1, \dots, \omega_n)}{q(\gamma) \mathbb{P}_{\theta_n|\gamma}(\omega_1, \dots, \omega_n)} - 1 = \frac{1}{q(\gamma)} \sum_{\alpha \neq \gamma} q(\alpha) \exp \left(\sum_{j \in \mathcal{A}} N_n(j) \ln \frac{p_{\theta_n}(j|\alpha)}{p_{\theta_n}(j|\gamma)} \right),$$

so that the result follows. \square

Now we state our main result.

Theorem 1 (LAMN). *Assume that **ID** holds and that for any $(j, \alpha) \in \mathcal{A} \times \mathcal{P}$ the function $p_{\cdot}(j|\alpha)$ is differentiable at θ^* . Then, the sequence $(\mathbb{P}_{\theta}|\mathcal{F}_n|\theta \in \Theta)$ is LAMN in θ^* with*

$$J_{\theta^*} := \sum_{j \in \mathcal{A}} p_{\theta^*}(j|\Gamma) \left(\nabla_{\theta} \ln p_{\theta}(j|\Gamma) (\nabla_{\theta} \ln p_{\theta}(j|\Gamma))^{\top} \right) \Big|_{\theta=\theta^*} =: I_{\theta^*}(\Gamma).$$

Proof. Fix $\gamma \in \mathcal{P}$. From standard results on parameter estimation for i.i.d. models, the sequence of models $(\mathbb{P}_{\theta|\gamma}|\mathcal{F}_n|\theta \in \Theta)$ is locally asymptotic normal in θ^* (see [14]). Namely, there exists a r.v. $Z^{\gamma} \sim \mathcal{N}(0, I_{\theta^*}(\gamma))$ such that, for any $h \in \Theta - \theta^*$,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \ln \frac{p_{\theta^* + \frac{h}{\sqrt{n}}}(\omega_k|\gamma)}{p_{\theta^*}(\omega_k|\gamma)} \stackrel{\mathcal{L} - \mathbb{P}_{\theta^*|\gamma}}{=} h^{\top} Z^{\gamma} - \frac{1}{2} h^{\top} I_{\theta^*}(\gamma) h.$$

From Lemma 2 setting $\theta_n = \theta^* + h/\sqrt{n}$

$$\lim_{n \rightarrow \infty} \ln \mathbb{P}_{\theta^* + \frac{h}{\sqrt{n}}}(\omega_1, \dots, \omega_n) - \ln \mathbb{P}_{\theta^* + \frac{h}{\sqrt{n}}|\gamma}(\omega_1, \dots, \omega_n) = -\ln q(\gamma), \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

and setting $\theta_n = \theta^*$,

$$\lim_{n \rightarrow \infty} \ln \mathbb{P}_{\theta^*}(\omega_1, \dots, \omega_n) - \ln \mathbb{P}_{\theta^*|\gamma}(\omega_1, \dots, \omega_n) = -\ln q(\gamma), \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

It follows that,

$$\lim_{n \rightarrow \infty} \ln \frac{\mathbb{P}_{\theta^* + \frac{h}{\sqrt{n}}}(\omega_1, \dots, \omega_n)}{\mathbb{P}_{\theta^*}(\omega_1, \dots, \omega_n)} - \sum_{k=1}^n \ln \frac{p_{\theta^* + \frac{h}{\sqrt{n}}}(\omega_k|\gamma)}{p_{\theta^*}(\omega_k|\gamma)} = 0, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Hence,

$$\lim_{n \rightarrow \infty} \ln \frac{\mathbb{P}_{\theta^* + \frac{h}{\sqrt{n}}}(\omega_1, \dots, \omega_n)}{\mathbb{P}_{\theta^*}(\omega_1, \dots, \omega_n)} \stackrel{\mathcal{L}-\mathbb{P}_{\theta^*|\gamma}}{=} h^\top Z^\gamma - \frac{1}{2} h^\top I_{\theta^*}(\gamma) h,$$

so that Lemma 1 yields the Theorem. □

4. Maximum likelihood estimation

This section is devoted to the study of the maximum likelihood estimator. For $\omega \in \Omega$ at step n the log likelihood is defined as

$$\ell_n(\theta)(\omega_1, \dots, \omega_n) = \frac{1}{n} \ln \mathbb{P}_\theta(\omega_1, \dots, \omega_n).$$

We study the maximum likelihood estimator defined as

$$\hat{\theta}_n := \operatorname{argmax}_{\theta \in \Theta} \ell_n(\theta).$$

4.1. Consistency

We define the function

$$\ell_{\theta^*, \gamma}(\theta) = -S_{\theta^*}(\gamma) - \min_{\alpha \in \mathcal{P}} S_{\theta^*|\theta}(\gamma|\alpha), \tag{6}$$

for all $\gamma \in \mathcal{P}$ and all $\theta \in \Theta$.

The following Lemma shows the almost sure uniform convergence of the sequence of log likelihood functions.

Lemma 3. *Assume that **ID** holds and that for any couple $(j, \alpha) \in \mathcal{A} \times \mathcal{P}$ the function $p.(j|\alpha)$ is continuous. Then for any $\gamma \in \mathcal{P}$,*

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\ell_n(\theta) - \ell_{\theta^*, \gamma}(\theta)| = 0, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.,$$

and $\ell_{\theta^*, \gamma}$ is such that, for any $\epsilon > 0$

$$\ell_{\theta^*, \gamma}(\theta^*) > \sup_{\theta: d(\theta, \theta^*) \geq \epsilon} \ell_{\theta^*, \gamma}(\theta). \tag{7}$$

Proof. Since Θ is compact and that for any $(j, \alpha) \in \mathcal{A} \times \mathcal{P}$ the function $p.(j|\alpha)$ is continuous and positive, we get

$$\max_{(j, \alpha) \in \mathcal{A} \times \mathcal{P}} \sup_{\theta \in \Theta} |\ln p_\theta(j|\alpha)| < \infty.$$

Furthermore, the strong law of large numbers (1) implies that for any $\alpha \in \mathcal{P}$,

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| \sum_{j \in \mathcal{A}} \frac{N_n(j)}{n} \ln \left(\frac{p_\theta(j|\alpha)}{p_{\theta^*}(j|\alpha)} \right) + S_{\theta^*|\theta}(\gamma|\alpha) \right| = 0, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Hence,

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| \frac{1}{n} \ln \frac{\mathbb{P}_{\theta|\alpha}(\omega_1, \dots, \omega_n)}{\mathbb{P}_{\theta^*|\gamma}(\omega_1, \dots, \omega_n)} + S_{\theta^*|\theta}(\gamma|\alpha) \right| = 0, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

It follows then from a repeated application of Lemma 5 (its statement and proof are postponed to the Appendix), that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\ell_n(\theta) - \ell_{\theta^*, \gamma}(\theta)| = 0, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Since $\ell_{\theta^*, \gamma}(\theta) = -S_{\theta^*}(\gamma) - \min_{\alpha \in \mathcal{P}} S_{\theta^*|\theta}(\gamma|\alpha)$, Assumption **ID** implies the inequality (7) and the lemma is proved. \square

The consistency of the maximum likelihood estimator follows now from standard arguments.

Proposition 2. *Assume that Assumption **ID** holds and that for any $(j, \alpha) \in \mathcal{A} \times \mathcal{P}$ the function $p.(j|\alpha)$ is continuous. Then,*

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta^*, \quad \mathbb{P}_{\theta^*} - a.s.$$

Proof. From Lemma 3, classical results of parametric estimation (see [14, Theorem 5.7]) give,

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta^*, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

So that, Lemma 1 yields the consistency w.r.t. \mathbb{P}_{θ^*} . \square

4.2. Saturation of Cramér–Rao bound

It is well know that if a model is LAMN, it verifies an asymptotic Cramér-Rao bound. Namely if (T_n) is a sequence of estimators such that

$$\lim_{n \rightarrow \infty} \sqrt{n}(T_n - (\theta^* + h/\sqrt{n})) \stackrel{\mathcal{L}-\mathbb{P}_{\theta^*+h/\sqrt{n}}}{=} T$$

for any $h \in \Theta - \theta^*$, then, $\mathbb{E}(T T^T) - \sum_{\alpha \in \mathcal{P}} q(\alpha) I_{\theta^*}^{-1}(\alpha)$ is positive semi definite [14, Corollary 9.9].

We now prove that the maximum likelihood estimator saturates this asymptotic bound. We prove it comparing $\hat{\theta}_n$ with $\hat{\theta}_n^\gamma$ defined by

$$\hat{\theta}_n^\gamma := \operatorname{argmax}_{\theta \in \Theta} \ell_n^\gamma(\theta),$$

where

$$\ell_n^\gamma(\theta)(\omega_1, \dots, \omega_n) = \frac{1}{n} \ln \mathbb{P}_{\theta|\gamma}(\omega_1, \dots, \omega_n)$$

for all $\theta \in \Theta$ and all $\omega \in \Omega$.

Lemma 4. Assume that **ID** holds and that for any $(j, \alpha) \in \mathcal{A} \times \mathcal{P}$ the function $p.(j|\alpha)$ is twice continuously differentiable in a neighborhood of θ^* . Assume further that for each $\gamma \in \mathcal{P}$, $I_{\theta^*}(\gamma)$ is not singular. Then,

$$\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^\gamma) = 0, \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Proof. Standard results of parametric estimation for i.i.d. random variables and our assumption **ID** imply $\lim_{n \rightarrow \infty} \hat{\theta}_n^\gamma = \theta^* \mathbb{P}_{\theta^*|\gamma}$ -a.s. From the definition of the maximum likelihood estimators,

$$\nabla \ell_n^\gamma(\hat{\theta}_n^\gamma) - \nabla \ell_n^\gamma(\hat{\theta}_n) = \nabla \ell_n(\hat{\theta}_n) - \nabla \ell_n^\gamma(\hat{\theta}_n). \tag{8}$$

Let \mathcal{O} be a sufficiently small neighborhood of θ^* (on which $p.$ is regular). The consistency of $\hat{\theta}_n$ and $\hat{\theta}_n^\gamma$ ensures that $\mathbb{P}_{\theta^*|\gamma}$ -a.s., for n large enough, the likelihood estimators $\hat{\theta}_n$ and $\hat{\theta}_n^\gamma$ belong to \mathcal{O} . Hence, since by assumption ℓ_n^γ is twice differentiable in a neighborhood of θ^* , there exists a sequence of random variable ξ_n lying in the segment with extremity $\hat{\theta}_n$ and $\hat{\theta}_n^\gamma$ such that $\mathbb{P}_{\theta^*|\gamma} - a.s.$, for n large enough,

$$\nabla^2 \ell_n^\gamma(\xi_n)(\hat{\theta}_n - \hat{\theta}_n^\gamma) = \nabla \ell_n^\gamma(\hat{\theta}_n) - \nabla \ell_n(\hat{\theta}_n).$$

The last equality comes from (8) and the Mean Value Theorem.

Now note that using explicit derivation and the strong law of large number (1), we have

$$\lim_{n \rightarrow \infty} \nabla^2 \ell_n^\gamma(\xi_n) = I_{\theta^*}(\gamma), \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Since $I_{\theta^*}(\gamma)$ is assumed to be non singular and $\theta \mapsto I_\theta(\gamma)$ is continuous in a neighborhood of θ^* , the previous convergence implies that for n large enough $\nabla^2 \ell_n^\gamma(\xi_n)$ is invertible. It follows that for n large enough,

$$|\hat{\theta}_n - \hat{\theta}_n^\gamma| = |[\nabla^2 \ell_n^\gamma(\xi_n)]^{-1}(\nabla \ell_n^\gamma(\hat{\theta}_n) - \nabla \ell_n(\hat{\theta}_n))| \quad \mathbb{P}_{\theta^*|\gamma} - a.s. \tag{9}$$

Let us now prove that $\nabla \ell_n^\gamma(\hat{\theta}_n) - \nabla \ell_n(\hat{\theta}_n) = o(n^{-1/2}) \mathbb{P}_{\theta^*|\gamma}$ -a.s.

Explicit differentiation leads to,

$$\begin{aligned} \nabla \ell_n^\gamma(\hat{\theta}_n) - \nabla \ell_n(\hat{\theta}_n) &= \sum_{\alpha \in \mathcal{P}} \sum_{j \in \mathcal{A}} \frac{N_n(j)}{n} \left(\frac{\nabla p_{\hat{\theta}_n}(j|\gamma)}{p_{\hat{\theta}_n}(j|\gamma)} - \frac{\nabla p_{\hat{\theta}_n}(j|\alpha)}{p_{\hat{\theta}_n}(j|\alpha)} \right) \\ &\quad \times \frac{q(\alpha) \mathbb{P}_{\hat{\theta}_n|\alpha}(\omega_1, \dots, \omega_n)}{\mathbb{P}_{\hat{\theta}_n}(\omega_1, \dots, \omega_n)}. \end{aligned}$$

From the consistency of $\hat{\theta}_n$ and the continuous differentiability of $p.(j|\alpha)$ in a neighborhood of θ^* , it follows that, there exists $C > 0$ such that for n large enough,

$$|\nabla \ell_n^\gamma(\hat{\theta}_n) - \nabla \ell_n(\hat{\theta}_n)| \leq C \left(1 - \frac{q(\gamma) \mathbb{P}_{\hat{\theta}_n|\gamma}(\omega_1, \dots, \omega_n)}{\mathbb{P}_{\hat{\theta}_n}(\omega_1, \dots, \omega_n)} \right) \quad \mathbb{P}_{\theta^*|\gamma} - a.s.$$

Equation (9) combined with the $\mathbb{P}_{\theta^*|\gamma}$ -a.s. convergence $\lim_{n \rightarrow \infty} \nabla^2 \ell_n^\gamma(\xi_n) = I_{\theta^*}(\gamma)$, and with the Lemma 2 yields the result. \square

Proposition 3. *Assume that **ID** holds and that for any $(j, \alpha) \in \mathcal{A} \times \mathcal{P}$ the function $p.(j|\alpha)$ is three times continuously differentiable in a neighborhood of θ^* . Assume further that for any $\gamma \in \mathcal{P}$, $I_{\theta^*}(\gamma)$ is not singular. Let $Z \sim \mathcal{N}(0, I_d)$ and Γ be a random variable independent of Z and taking value in \mathcal{P} such that $\Pr(\Gamma = \gamma) = q(\gamma)$. Then, for any $h \in \Theta - \theta^*$,*

$$\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\theta}_n - (\theta^* + h/\sqrt{n})) \stackrel{\mathcal{L}-\mathbb{P}_{\theta^*+h/\sqrt{n}}}{=} I_{\theta^*}(\Gamma)^{-\frac{1}{2}} Z.$$

Proof. From standard results in parameter estimation[14], under Assumption **ID** the statistical model $(\mathbb{P}_{\theta|\gamma}, \theta \in \Theta)$ is LAN. Moreover, for any $h \in \Theta - \theta^*$, under the assumptions of the proposition,

$$\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\theta}_n^\gamma - (\theta^* + h/\sqrt{n})) \stackrel{\mathcal{L}-\mathbb{P}_{\theta^*|\gamma}}{=} I_{\theta^*}(\gamma)^{-\frac{1}{2}} Z - h. \tag{10}$$

Actually, the proof of $(\mathbb{P}_{\theta|\gamma}, \theta \in \Theta)$ being LAN and the weak convergence (10) are based on the same Central Limit Theorem. It follows that,

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\sqrt{n}(\hat{\theta}_n^\gamma - (\theta^* + h/\sqrt{n})), \ln \frac{\mathbb{P}_{\theta^*+h/\sqrt{n}|\gamma}(\omega_1, \dots, \omega_n)}{\mathbb{P}_{\theta^*|\gamma}(\omega_1, \dots, \omega_n)} \right) \\ \stackrel{\mathcal{L}-\mathbb{P}_{\theta^*|\gamma}}{=} \left(I_{\theta^*}^{-\frac{1}{2}}(\gamma)Z - h, h^\top I_{\theta^*}(\gamma)^{\frac{1}{2}}Z - \frac{1}{2}h^\top I_{\theta^*}(\gamma)h \right). \end{aligned}$$

It follows then from Lemmas 2 and 4 that,

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\sqrt{n}(\hat{\theta}_n - (\theta^* + h/\sqrt{n})), \ln \frac{\mathbb{P}_{\theta^*+h/\sqrt{n}}(\omega_1, \dots, \omega_n)}{\mathbb{P}_{\theta^*}(\omega_1, \dots, \omega_n)} \right) \\ \stackrel{\mathcal{L}-\mathbb{P}_{\theta^*|\gamma}}{=} \left(I_{\theta^*}^{-\frac{1}{2}}(\gamma)Z - h, h^\top I_{\theta^*}(\gamma)^{\frac{1}{2}}Z - \frac{1}{2}h^\top I_{\theta^*}(\gamma)h \right). \end{aligned}$$

Now, from Le Cam's third Lemma we get

$$\lim_{n \rightarrow \infty} \sqrt{n}(\hat{\theta}_n - (\theta^* + h/\sqrt{n})) \stackrel{\mathcal{L}-\mathbb{P}_{\theta^*+h/\sqrt{n}}}{=} I_{\theta^*}(\gamma)^{-\frac{1}{2}} Z.$$

So that, Lemma 1 yields the proposition. □

5. Applications to quantum non-demolition measurement

As mentioned in the Introduction, the above development is motivated by some applications in quantum physics. In particular, as we will see, the above estimation results can be applied in the context of QND measurement. For the sake of completeness we recall briefly the QND model. For a complete overview of this model we refer to [2, 3].

Let $\{e_\alpha \mid \alpha \in \mathcal{P}\}$ and $\{\psi_j \mid j \in \mathcal{A}\}$ be orthonormal basis of respectively \mathbb{C}^d and \mathbb{C}^l . These last spaces are endowed with their canonical Hilbert space structure. In the context of quantum physics, these basis will be associated with

some physical quantities. Each vector of these basis describes the physical state corresponding to an almost sure value of said physical quantities. The Hilbert space \mathbb{C}^d describes the quantum system that one aims to measure indirectly. The Hilbert space \mathbb{C}^l describes a probe that will be used to measure indirectly. Now, we detail the usual setup of indirect measurement. It consists in measuring something on the probe after some interaction between the system and the probe. More precisely, the interaction is described through a unitary operator U on $\mathbb{C}^d \otimes \mathbb{C}^l$. For QND measurements, this operator may be written as

$$U = \sum_{\alpha \in \mathcal{P}} \pi_{e_\alpha} \otimes U_\alpha.$$

Here, π_{e_α} is the projector on the line $\mathbb{C}e_\alpha$ and $(U_\alpha)_{\alpha \in \mathcal{P}}$ are unitary operators on \mathbb{C}^l . (U_α) depends on the unknown parameters of the experiment. The state of the system is represented by the unit vector¹ $\phi_0 \in \mathbb{C}^d$. This vector may be expanded on the first basis:

$$\phi_0 = \sum_{\alpha} \langle e_\alpha, \phi_0 \rangle e_\alpha.$$

The state of the probe is represented by a unit vector $\psi \in \mathbb{C}^l$. After the interaction the joint system–probe state is the unit vector $U(\phi_0 \otimes \psi) \in \mathbb{C}^d \otimes \mathbb{C}^l$. This vector may be expanded as,

$$U(\phi_0 \otimes \psi) = \sum_{\alpha} \langle e_\alpha, \phi_0 \rangle e_\alpha \otimes U_\alpha \psi.$$

We are now in position to see how multinomial mixtures encompass the law of sequence of measurement results in QND measurements.

To begin with, let us assume that $\phi_0 = e_\alpha$ for some $\alpha \in \mathcal{P}$. Then,

$$q_0(\alpha) := |\langle e_\alpha, \phi_0 \rangle|^2 = 1.$$

Further, from the definition of U ,

$$U(\phi_0 \otimes \psi) = e_\alpha \otimes U_\alpha \psi.$$

This property justifies the denomination *non demolition* explained in the Introduction. If the system state is e_α before the interaction it remains e_α after the interaction. Quantum mechanics tells that measuring a physical state in $\{\psi_j \mid j \in \mathcal{A}\}$ has the following probability distribution

$$\mathbb{P}[\text{observing } \psi_j] = |\langle \psi_j, U_\alpha \psi \rangle|^2.$$

We set

$$p(j|\alpha) := |\langle \psi_j, U_\alpha \psi \rangle|^2.$$

¹Actually the state corresponds to the line given by the direction ϕ_0 . Particularly two states are equivalent if they differ only by a phase. We no longer mention it and always mean equality up to a phase when comparing two states.

We will later see how U_α depends on the unknown parameter θ (and will add an index θ to this notation). If $\phi_0 \notin \{e_\alpha \mid \alpha \in \mathcal{P}\}$, the probe measurement outcome is ψ_j with probability

$$\begin{aligned}\pi_0(j) &= \mathbb{P}[\text{observing } \psi_j] = \sum_{\alpha} |\langle e_\alpha, \phi_0 \rangle|^2 |\langle \psi_j, U_\alpha \psi \rangle|^2 \\ &= \sum_{\alpha} q_0(\alpha) p(j|\alpha).\end{aligned}$$

The quantum mechanics projection postulate implies that if ψ_j is the measurement outcome then the joint system–probe state becomes

$$\tilde{\phi}_1(j) = \frac{\sum_{\alpha} \langle \phi_0, \alpha \rangle \langle U_\alpha \psi, j \rangle}{\sqrt{\pi_0(j)}} \otimes \psi_j.$$

Hence, the system state goes from ϕ_0 to

$$\phi_1(j) = \frac{\sum_{\alpha} \langle \phi_0, \alpha \rangle \langle U_\alpha \psi, j \rangle}{\sqrt{\pi_0(j)}}. \quad (11)$$

In others words, this is the new system state conditioned on the outcome ψ_j (for the probe). This system state update leads to the update of $q_0(\alpha)$,

$$q_1(\alpha) := |\langle e_\alpha, \phi_1 \rangle|^2 = q_0(\alpha) \frac{p(j|\alpha)}{\pi_0(j)}.$$

This procedure results in the definition of random variables ϕ_1 and $(q_1(\alpha))_{\alpha \in \mathcal{P}}$ whose laws are images of the law $\mathbb{P}[\text{observing } \psi_j] = \pi_0(j)$.

Now, we repeat the previous steps. Let (X_n) be the resulting sequence of outcome (identifying ψ_j and j). We have

$$\mathbb{P}[X_1 = j] = \pi_0(j) = \sum_{\alpha} q_0(\alpha) p(j|\alpha), \quad (j \in \mathcal{A}), \quad (12)$$

$$\mathbb{P}[X_1 = j, X_2 = j'] = \sum_{\alpha} q_0(\alpha) p(j|\alpha) p(j'|\alpha), \quad (j, j' \in \mathcal{A}), \quad (13)$$

$$\mathbb{P}[X_1 = j_1, \dots, X_n = j_n] = \sum_{\alpha} q_0(\alpha) \prod_{k=1}^n p(j_k|\alpha), \quad (n \in \mathbb{N}_*; j_1, \dots, j_n \in \mathcal{A}). \quad (14)$$

Using Kolomgorov's consistency Theorem we thus have defined the law of the random sequence (X_n) . For $\alpha \in \mathcal{P}$, let \mathbb{P}_α be the probability measure such that

$$\mathbb{P}_\alpha[\{(j_1, \dots, j_n, \omega) : \omega \in \Omega\}] = \prod_{k=1}^n p(j_k|\alpha). \quad (15)$$

Then the law of (X_n) is the mixture of multinomials $\sum_{\alpha} q_0(\alpha) \mathbb{P}_\alpha$. Let turn now to the statistical model. For any $\alpha \in \mathcal{P}$, the unitary operator $U_\alpha = U_\alpha(\theta)$

depends on the unknown parameter θ . Hence, we wish to study the statistical model (\mathbb{P}_θ) with

$$\mathbb{P}_\theta := \sum_{\alpha} q_0(\alpha) \mathbb{P}_{\theta, \alpha},$$

where $\mathbb{P}_{\theta, \alpha}$ is defined in (15) with

$$p_\theta(j|\alpha) = |\langle U_\alpha(\theta)\psi, j \rangle|^2.$$

Now, all the results developed in the last sections hold whenever the regularity assumptions are assumed directly on $(U_\alpha(\cdot))_{\alpha \in \mathcal{P}}$. Let us now unravel what is the Fisher information for the QND measurement model.

Let $\alpha \in \mathcal{P}$ and Θ be some given non empty open subset of an Euclidean space. Further, let $H_\alpha(\cdot)$ be a differentiable function on Θ taking its values in the set of self-adjoint $l \times l$ matrices. For $\alpha \in \mathcal{P}$, set $U_\alpha(\cdot) := \exp(-iH_\alpha(\cdot))$. Then, a direct differentiation gives,

$$\partial_{\theta_k} \ln(p(i|\alpha)) = 2 \operatorname{Im} \left(\frac{\langle \partial_{\theta_k} H_\alpha(\cdot) U_\alpha(\cdot) \psi, \psi_j \rangle}{\langle U_\alpha(\cdot) \psi, \psi_j \rangle} \right).$$

It follows that,

$$\begin{aligned} (I(\alpha))_{kl} &= 4 \sum_{j \in \mathcal{A}} p(j|\alpha) \operatorname{Im} \left(\frac{\langle \partial_{\theta_k} H_\alpha(\cdot) U_\alpha(\cdot) \psi, \psi_j \rangle}{\langle U_\alpha(\cdot) \psi, \psi_j \rangle} \right) \\ &\quad \times \operatorname{Im} \left(\frac{\langle \partial_{\theta_l} H_\alpha(\cdot) U_\alpha(\cdot) \psi, \psi_j \rangle}{\langle U_\alpha(\cdot) \psi, \psi_j \rangle} \right). \end{aligned}$$

More particularly, in the relevant one dimensional case $d = 1$ where $H_\alpha(\theta) = \theta H_\alpha$ for some self adjoint matrices H_α ,

$$\partial_\theta \ln(p_\theta(i|\alpha)) = 2 \operatorname{Im} \left(\frac{\langle H_\alpha U_\alpha(\theta) \psi, \psi_j \rangle}{\langle U_\alpha(\theta) \psi, \psi_j \rangle} \right).$$

So that,

$$I_\theta(\alpha) = 4 \sum_{j \in \mathcal{A}} p_\theta(j|\alpha) \operatorname{Im}^2 \left(\frac{\langle H_\alpha U_\alpha(\theta) \psi, \psi_j \rangle}{\langle U_\alpha(\theta) \psi, \psi_j \rangle} \right).$$

Toy QND example:

We conclude our paper by an illustration of our results on the Haroche’s group experiment [8] mentionned in the Introduction. In that experiment $d = 8$ and $\mathcal{A} = \{0, 1\} \times \{0, 1, 2, 3\}$ hence $l = 8$ and j is defined through a bijection $j \equiv (x, a)$. The expression of $p(j|\alpha)$ is provided in [8]. Expressed in our notations we have here,²

$$p_\theta(x, a|\alpha) = (\theta_5 + \theta_6 \cos(\alpha\theta_4 + \theta_a + x\pi))/8.$$

²Remark that in [8] $\alpha = 0, \dots, 7$ so Assumption **ID** is not verified for $\alpha = 0$. More precisely θ_4 may not be identifiable. Though for the ideal value of $\theta = \pi/4$, $\alpha = 0$ is equivalent to $\alpha = 8$ so we use this value. Hence, we identify the zero photon state with the eight one at the opposite of what is done in [8].

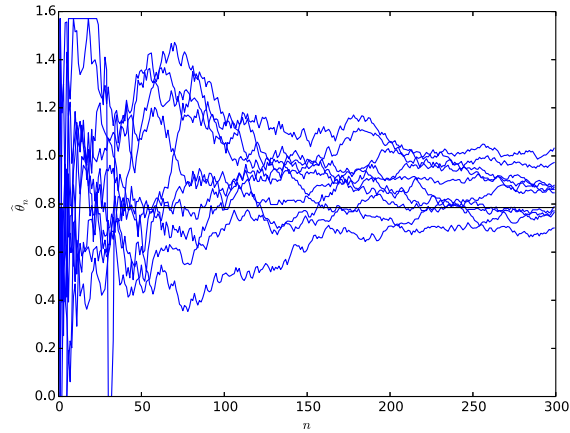


FIG 1. Numerical simulation of 10 realizations of (θ_n) for the toy model described in the text with $q(\alpha)$ proportional to $3.46^\alpha/\alpha!$ for $\alpha \in \{1, \dots, 8\}$. The black line corresponds to the target value $\theta^* = \pi/4$.

Hence, $\dim \Theta = 7$. The ideal values of the parameters are $\theta_5 = \theta_6 = 1$, $\theta_4 = \pi/4$ and $\theta_a = (2-a)\pi/4$. Experimentally some imperfections imply that θ_6 is smaller than 1. It is close to $\theta = 0.674 \pm 0.004$. Besides this limitation in [8] the authors find parameters close to their target values using a best fit to the empirical distribution. Setting $\theta_6 = 0.674$, it is easy to check that **ID** holds for the ideal parameters and in a small enough but sufficiently large neighborhood of the true parameters. Moreover, all the functions $\theta \mapsto p_\theta(x, a|\alpha)$ are entire analytic. We limit ourselves to the estimation of θ_4 the other parameters are fixed to their true values except θ_6 . We set $\theta_6 = 0.674$. Further, we take $\Theta = [\pi/8, 3\pi/8]$ and,

$$p_\theta(x, a|\alpha) := (1 + 0.674 \cos(\alpha\theta + (2 - a)\pi/4 + x\pi))/8.$$

It follows that

$$I_\theta(\alpha) = \sum_{a=0}^3 \alpha^2 \frac{0.674^2}{4} \frac{\sin^2(\alpha\theta + (2 - a)\pi/4)}{1 - 0.674^2 \cos^2(\alpha\theta + (2 - a)\pi/4)}.$$

Fisher information is not singular at $\theta = \pi/4$. Figure 1 depicts some simulations of $(\hat{\theta}_n)$ for this toy model.

Appendix A: Uniform convergence of logarithm of sums of uniformly convergent sequences

Lemma 5. Let (a_n) and (b_n) be two sequence of strictly positive functions with common domain of definition such that there exists two functions ℓ_a and ℓ_b such that,

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \ln a_n - \ell_a \right\|_\infty = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \left\| \frac{1}{n} \ln b_n - \ell_b \right\|_\infty = 0$$

with $\|\cdot\|_\infty$ the sup norm. Then,

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \ln(a_n + b_n) - \max(\ell_a, \ell_b) \right\|_\infty = 0.$$

Proof. Let D be the common domain of definition of the functions. Assumptions imply there exists a sequence (ϵ_n) of strictly positive numbers such that $\lim_{n \rightarrow \infty} \epsilon_n = 0$ and

$$\max \left(\left\| \frac{1}{n} \ln a_n - \ell_a \right\|_\infty, \left\| \frac{1}{n} \ln b_n - \ell_b \right\|_\infty \right) \leq \epsilon_n.$$

Since the function \ln is non decreasing, we deduce first that for any $x \in D$,

$$\frac{1}{n} \ln(a_n(x) + b_n(x)) - \max(\ell_a(x), \ell_b(x)) \leq \epsilon_n + \frac{1}{n} \ln 2.$$

Second, let $c(x)$ be such that $\ell_c = \max(\ell_a, \ell_b)$. Then for any $x \in D$,

$$-\epsilon_n \leq \frac{1}{n} \ln c_n(x) - \ell_c(x) \leq \frac{1}{n} \ln(a_n(x) + b_n(x)) - \max(\ell_a(x), \ell_b(x))$$

and the Lemma holds. \square

Acknowledgments

We thank the referee for useful comments. T.B and C.P are supported by ANR project StoQ ANR-14-CE25-0003-01 and CNRS InFiniTi project MISTEQ. The research of T.B. has been supported by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02.

References

- [1] M. Ballesteros, M. Fraas, J. Fröhlich and B. Schubnel, *Indirect acquisition of information in quantum mechanics*. J. Stat. Phys. **162** (2016) 924–958. [MR3456983](#)
- [2] M. Bauer and D. Bernard, *Convergence of repeated quantum nondemolition measurements and wave-function collapse*. Phys. Rev. A **84** (2011) 044103
- [3] M. Bauer, T. Benoist and D. Bernard, *Repeated quantum non-demolition measurements: convergence and continuous time limit*. Ann. Henri Poincaré **14** (2013) 639–679. [MR3046451](#)
- [4] M. Bauer, D. Bernard and T. Benoist, *Iterated stochastic measurements*. J. Phys. A **45** (2012) 494020. [MR2999738](#)
- [5] T. Benoist and C. Pellegrini, *Large time behavior and convergence rate for quantum filters under standard non demolition conditions*. Comm. Math. Phys. **331** (2014) 703–723. [MR3238528](#)
- [6] V. B. Braginsky, YI. Vorontsov and K. S. Thorne. *Quantum nondemolition measurements*. Science, **209**(4456) (1980) 547–57.
- [7] C. Catana, L. Bouten and M. Guță, *Fisher informations and local asymptotic normality for continuous-time quantum Markov processes*. J. Phys. A **48** (2015) 365301. [MR3400896](#)

- [8] C. Guerlin, J. Bernu, S. Deleglise, C. Sayrin, S. Gleyzes, S. Kuhr, M. Brune, J. M. Raimond and S. Haroche, *Progressive field-state collapse and quantum non-demolition photon counting* Nature **448** (2007) 889–893.
- [9] M. Guță and A. Jenčová, *Local asymptotic normality in quantum statistics* Commun. Math. Phys., **276** (2007) 34–379. [MR2346393](#)
- [10] M. Guță and J. Kiukas, *Equivalence classes and local asymptotic normality in system identification for quantum Markov chains*. arXiv preprint arXiv:1402.3535 (2014). [MR3320317](#)
- [11] A. S. Holevo, *Statistical structure of quantum theory*. Springer, 2003. [MR1889193](#)
- [12] J. Kahn and M. Guță, *Local asymptotic normality and optimal estimation for finite dimensional quantum systems* in Quantum stochasticity and information: Statistics, Filtering and Control, V. P. Belavkin and M. Guță editors, World Scientific 2008. [MR2562438](#)
- [13] P. Six, P. Campagne-Ibarcq, L. Bretheau, B. Huard and P. Rouchon, *Parameter estimation from measurements along quantum trajectories*, Proc. IEEE Conf. on Decision and Control, (2015) 7742–7748.
- [14] A. W. van der Vaart, *Asymptotic Statistics* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press 1998 [MR1652247](#)