

# The Quantum State Estimation (introduction)

Keiji Matsumoto

ERATO Project Quantum computation and Information,  
Japan Science and Technology

keiji@qci.jst.go.jp

## Contents

1. What is quantum state estimation ?  
problem, motivation
2. Classical Statistics  
problem  
baysian  
asymptotic theory
3. Quantum First Order Asymptotic Theory  
formalim,  
Collective measurement and Separable measurement  
Single letterization theorem  
An application

## Classical Statistics

estimate unknown probability distribution from data

$$x_1, x_2, \dots, x_n \sim p(x, \theta), \\ \theta \in \Theta \subset \mathbb{R}^m$$

estimate:  $\hat{\theta}(x_1, x_2, \dots, x_n)$

### Example 1

$$p(x, \theta) = \frac{1}{\sqrt{2}} \exp -\frac{1}{2}(x - \theta)^2$$

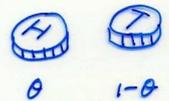
$$\hat{\theta}_1(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{\theta}_2(x_1, \dots, x_n) = \frac{1}{n/2} \sum_{i=1}^{n/2} x_i$$

$$\hat{\theta}_3(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \alpha_i x_i$$

etc

### Example 2



$$p(x, \theta) = \begin{cases} \theta & x = 0 \\ 1 - \theta & x = 1 \end{cases} \\ 0 \leq \theta \leq 1$$

$$\hat{\theta}_1(x_1, \dots, x_n) = \frac{1}{n} (\text{the number of 0s})$$

$$\hat{\theta}_2(x_1, \dots, x_n) = \frac{1}{n} (\text{the number of 0s})$$

etc  $\hat{\theta}_3 = \frac{1}{2}$

- Which estimate is optimal ?
- How to measure goodness/badness of the estimate ?

**measure of the efficiency**

Cost function :  $d(\hat{\theta}, \theta)$ , (The Smaller, the better)

For example :

$$\sum_i g_i(\hat{\theta}^i - \theta)^2$$

$$\cos(\|\hat{\theta} - \theta\|)$$

etc

• The smaller the cost is, the better.

→ measure of badness ?

• But,  $\hat{\theta}$  is probabilistic !

→ take average

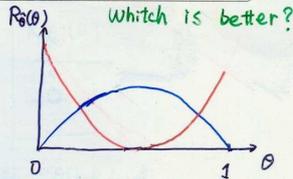
$R_{\hat{\theta}(\cdot)}(\theta) :=$

$$\int \|\hat{\theta}(x_1, \dots, x_n) - \theta\|^2 p(x_1, \theta) \dots p(x_n, \theta) dx_1 \dots dx_n$$

→ measure of badness

$R(\theta)$ , the average of the cost, is still dependent on  $\theta$ .

|                                                       | $R_{\hat{\theta}}(\theta)$     |
|-------------------------------------------------------|--------------------------------|
| $\hat{\theta}_1 = \frac{1}{n}(\# \text{ of } \oplus)$ | $\frac{1}{n} \theta(1-\theta)$ |
| $\hat{\theta}_2 = \frac{1}{2}$                        | $(\theta - \frac{1}{2})^2$     |



**Basyan approach**

Assume prior distribution  $q(\theta)$  of  $\theta \in \Theta$

Badness of the estimator :  $\int R_{\hat{\theta}(\cdot)}(\theta)q(\theta)d\theta$

Finding best estimate  $\Leftrightarrow$

minimize  $\int R_{\hat{\theta}(\cdot)}(\theta)q(\theta)d\theta$

• merit – mathematically simple

• demerit –

– how to chose prior?

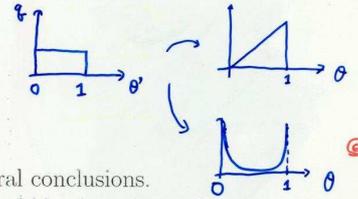
\* uniform distribution on  $\Theta$ ?

→ But 'uniform distribution' depends on which coordinate you take.

In Example 2...

(1)  $\theta' = (\theta)^2$

(2)  $\theta' = \cos^{-1} \sqrt{\theta}$



– hard to obtain general conclusions.

the minimization should be down for each problems.

**Sampling Theory**

• do not assume prior distribution.

• compare  $R_{\hat{\theta}(\cdot)}(\theta)$  itself.

• impose some conditions on  $\hat{\theta}(\cdot)$ , to exclude estimate like  $\hat{\theta}_3(\cdot)$ .

• merit – result is very persuasive, if obtained.

• demerit – hard to obtain clear result

**Sampling Theory, Asymptotic Setting**

• do not assume prior distribution

• but assume number of data is large

• merit

– quite general result is obtained

**Asymptotic Cramér-Rao Inequality**

If  $V_{\theta}[\hat{\theta}] \rightarrow 0$  as  $n \rightarrow \infty$ ,

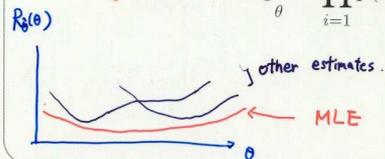
$$V_{\theta}[\hat{\theta}] \geq \frac{1}{n} J_{\theta}^{-1} + o\left(\frac{1}{n}\right). \quad (1)$$

\*  $J_{\theta}$  : Fisher Information Matrix

$$[J_{\theta}]_{i,j} := E_{\theta}[\partial_i \log p(x_{\theta}) \partial_j \log p(x_{\theta})]$$

\* equality is achieved by Maximum Likelihood Estimator at any  $\theta$ !

$$\hat{\theta}_{MLE} := \operatorname{argmax}_{\theta} \prod_{i=1}^n p(x_i, \theta)$$

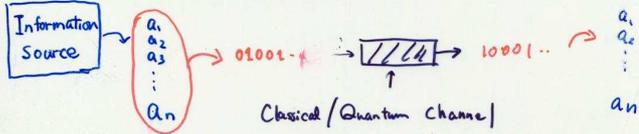


– optimal estimate (MLE) is practically useful.

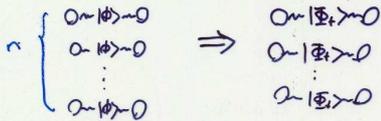
• demerit – if number of data is small ...

**Asymptotic Theories in other fields of Classical/Quantum Information Sciences**

- Classical/Quantum Shannon Theory
  - Shannon Theorem (Classical)
  - Holevo's Theorem (Classical)



- Entanglement Concentration/Distillation/Dilution



- Merit of asymptotic theory
  - quite general result is obtained
  - sometimes, asymptotically optimal methods are practical. (classical statistical estimation, classical data compression)
- Demerit of asymptotic theory - if  $n$  is not large

**Single Letterization Theorem**

(Nagaoka 1989, Hayashi and Matsumoto 1998, Gill and Massar 2000)

Minimum of  $A_\rho$  over separable measurements in  $\mathcal{H}^{\otimes n}$   
 ( $\equiv C_\theta$ , Quasi-quantum Cramér-Rao bound)

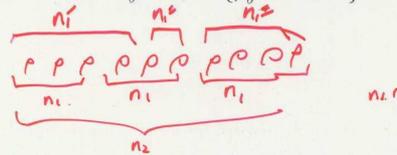
$$\min_{M: \text{measurement in } \mathcal{H}} \text{Sp } J_\theta^S [J_\theta^M]^{-1}$$

- Sp : trace over  $\mathbf{R}^m$
- $J_\theta^S$  : SLD Fisher Information Matrix
- $J_\theta^M$  : Classical Fisher Information Matrix by Measurement  $M$

**Limiting Expression**

(Hayashi and Matsumoto 1998)

- $C_\theta^Q = \lim_{n \rightarrow \infty} n C_\theta^n$ , where  $C_\theta^n : C_\theta$  of  $\{\rho_\theta^{\otimes n} : \theta \in \Theta\}$



**Quantum Asymptotic Theory**

**Basic Formalism**

- Given  $n$  copies  $\rho^{\otimes n}$  of unknown states  $\rho$ 
  - $n$  is very large ( $n \gtrsim 10^2$ )

$$E_\rho[1 - F(\rho, \hat{\rho}_n)] = A_\rho \frac{1}{n} + o\left(\frac{1}{n}\right) \quad (2)$$

- $F(\cdot, \cdot)$  : Fidelity
- $\hat{\rho}_n$  estimated state given  $n$  copies

- Optimize the measurement and the data processing to minimize  $A_\rho$

**Statistical Model**

Often, unknown state is limited to a member of a set

$$\mathcal{M} = \{\rho_\theta ; \theta \in \Theta \in \mathbf{R}^m\}$$

**Example 3**

$$|\phi_\theta\rangle = e^{-i(\theta^1 P - \theta^2 Q)} |\phi_0\rangle,$$

- $|\phi_0\rangle$  : known
- $P$  : Momentum
- $Q$  : Position

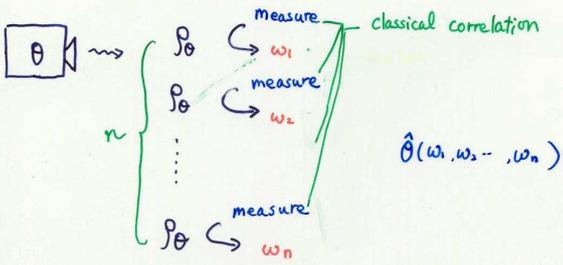
**Example 4**  $\rho_\theta$  is a thermal gaussian state,

$$\rho_\theta = c_{\theta^3} e^{-i(\theta^1 P - \theta^2 Q)} e^{-\theta^3 N} e^{i(\theta^1 P - \theta^2 Q)}$$

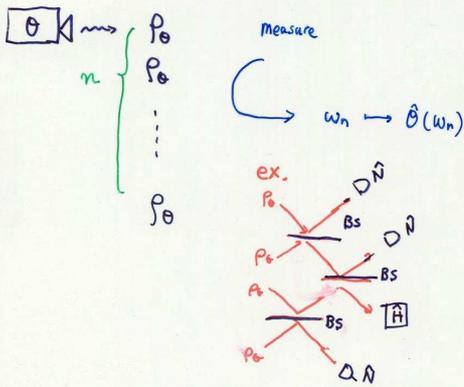
- $N$  : Number operator
- $c_{\theta^3}$  : constant

### Separable and Collective Measurement (I)

#### Separable Measurement



#### Collective measurement



### Separable and Collective Measurement (II)

$$C_\theta^Q \equiv \lim_{n \rightarrow \infty} \min_{\text{all measurements in } \mathcal{H}^{\otimes n}} n E_\theta [1 - F(\rho_\theta, \rho_{\hat{\theta}_n})]$$

- $C^Q \leq C$  (by definition)
- $C^Q = C \Leftrightarrow$  Collective measurement is NOT effective

- pure state models (Matsumoto, 2002)
- one parameter models (Nagaoaka, 1990)

$C^Q < C \Leftrightarrow$  Collective measurement IS effective

- $\mathcal{M}$  is the totality of all the mixed states

- \* 2-level system : Hayashi 1998, Gill and Massar 2000
- \* d-level system : Matsumoto

- Thermal State Model (example 4) (Hayashi 1998)

- etc...

$$C_\theta \equiv \lim_{n \rightarrow \infty} \min_{\text{separable measurements}} n \cdot E_\theta [1 - F(\rho_\theta, \rho_{\hat{\theta}_n})]$$