Asymptotic methods in Quantum Statistics

Mădălin Guță

School of Mathematics

University of Nottingham

EPSRC

Modern statistical methods in QIP

Heraeus summer school, Bad Honnef 2011

The University of Nottingham

Outline of the mini-course

1. Notions of statistical inference
2. Basics of quantum state estimation
3. The 8 (14) ions experiment
4. The quantum Cramér-Rao theory

Quantum Statistics

5. Local asymptotic normality for i.i.d. quantum states

System identification for quantum Markov processes

What this course does not cover (but is worth knowing)

- Bayesian methods
- Covariant estimation methods
- Channel/phase estimation
- Compressed sensing
- Quantum Homodyne Tomography
- Quantum Metrology
- Quantum cloning, teleportation benchmarks, learning ...

A short and biased list of references

Classical Statistics

- G.A. Young and R.L. Smith, *Essentials of statistical inference*, Cambridge Univ. Press (2005)
- D.R. Cox, *Principles of Statistical Inference*, Cambridge University Press (2006)
- A. van der Vaart, *Asymptotic Statistics*, Cambridge University Press (2000)

Quantum Statistics

- C. W. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press (1976)
- A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* North Holland (1982) (second edition in Publications of the Scuola Normale Superiore (2011))
- M. Hayashi (editor), *Asymptotic theory of quantum statistical inference*, World Scientific (2005)
- M. Paris and J. Rehacek, (editors) *Quantum State Estimation*, Springer (2008)
- O. E. Barndorff-Nielsen, R.D. Gill, P.E. Jupp, On quantum statistical inference, *J. R. Stat. Soc. B Stat. Methodol.* 65, 775-816 (2003)
- A. I. Lvovsky and M. G. Raymer, Continuous-variable optical quantum-state tomography *Reviews of Modern Physics* 81, 299-332 (2009)
- M. Guţă, *Quantum Statistics*, 10 hours course at http://maths.dept.shef.ac.uk/magic/course.php?id=181

1. Notions of statistical inference

- Statistical models
- Parametric estimation
- Fisher Information
- Cramér-Rao bound
- Efficient estimators
- Repeated coin toss example
- Local asymptotic normality
- Confidence intervals and Bootstrap
- Hypothesis testing

What is statistical inference?

Given random data X drawn from an unknown distribution, one aims to make an 'educated guess' about some property of the underlying distribution

Example

- Density estimation: given $X_1, \ldots, X_n$ independent identically distributed (i.i.d.) with unknown density $p \in L^1([0,1])$, estimate the value of $p(x)$ for some $x \in [0,1]$
- Hypothesis testing: given X drawn from either $\mathbb{P}_0$ or $\mathbb{P}_1$ decide from which of the two distributions it comes
- Sufficient statistic: can data $X \sim \mathbb{P}_\theta$ be 'summarised' into a 'simpler' statistics $f(X)$ without losing information about θ ?
- Identifiability: Is the map $\theta \mapsto \mathbb{P}_\theta$ one-to-one ?
- Optimality: how do we compare the performance of estimators and which are the optimal ones?
- Asymptotics: what happens in the limit of 'large number of data'?

Statistical models

Definition

Let Θ be a parameter space. A statistical model over Θ is a family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of probability distributions on a measure space $(\mathcal{X}, \Sigma)$.

Definition

Let Θ be a parameter space. A statistical model over Θ is a family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of probability distributions on a measure space $(\mathcal{X}, \Sigma)$.

Example

- Repeated coin toss: $X_1, \ldots, X_n$ i.i.d. with $\mathbb{P}_\theta([X_i = 1]) = \theta$ and $\mathbb{P}_\theta([X_i = 0]) = 1 - \theta$, with $\theta \in \Theta := [0, 1]$. The joint distribution is:

$$\mathbb{P}^n_\theta([X_1 = x_1, \ldots, X_n = x_n]) = \prod_{i=1}^{n} \mathbb{P}_\theta([X_i = x_i]) = \theta^{\sum_i x_i} \cdot (1-\theta)^{n - \sum_i x_i}$$

Statistical models

Definition

Let Θ be a parameter space. A statistical model over Θ is a family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of probability distributions on a measure space $(\mathcal{X}, \Sigma)$.

Example

- Repeated coin toss: $X_1, \ldots, X_n$ i.i.d. with $\mathbb{P}_\theta([X_i = 1]) = \theta$ and $\mathbb{P}_\theta([X_i = 0]) = 1 - \theta$, with $\theta \in \Theta := [0, 1]$. The joint distribution is:

$$\mathbb{P}^n_\theta([X_1 = x_1, \ldots, X_n = x_n]) = \prod_{i=1}^{n} \mathbb{P}_\theta([X_i = x_i]) = \theta^{\sum_i x_i} \cdot (1-\theta)^{n - \sum_i x_i}$$

- Gaussian shift on $\mathbb{R}^k$: family of Gaussian distributions $N(\theta, V)$ with unknown mean $\theta \in \mathbb{R}^k$ and known $k \times k$ covariance matrix V

Definition

Let Θ be a parameter space. A statistical model over Θ is a family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of probability distributions on a measure space $(\mathcal{X}, \Sigma)$.

Example

- Repeated coin toss: $X_1, \ldots, X_n$ i.i.d. with $\mathbb{P}_\theta([X_i = 1]) = \theta$ and $\mathbb{P}_\theta([X_i = 0]) = 1 - \theta$, with $\theta \in \Theta := [0, 1]$. The joint distribution is:

$$\mathbb{P}_\theta^n([X_1 = x_1, \ldots, X_n = x_n]) = \prod_{i=1}^{n} \mathbb{P}_\theta([X_i = x_i]) = \theta^{\sum_i x_i} \cdot (1 - \theta)^{n - \sum_i x_i}$$

- Gaussian shift on $\mathbb{R}^k$: family of Gaussian distributions $N(\theta, V)$ with unknown mean $\theta \in \mathbb{R}^k$ and known $k \times k$ covariance matrix V

- Tomography: an unknown probability density p over $\mathbb{R}^2$ is probed through its marginals along random directions ϕ in plane. For each ϕ we get data $X \sim \mathcal{R}[p](x|\phi)$ where $\mathcal{R}[p]$ is the Radon transform

$$\mathcal{R}[p](x|\phi) = \int p(x \cos\phi + t \sin\phi, x \sin\phi - t \cos\phi) dt$$

Problem

Given

- *a subset* Θ *of* $\mathbb{R}^k$
- *data* $X \sim \mathbb{P}_\theta$ *with* $\theta \in \Theta$ *and* $\mathbb{P}_\theta$ *probability distribution on* $(\mathcal{X}, \Sigma)$
- *a loss function* $W : \Theta \times \Theta \to \mathbb{R}_+$, *e.g.* $W(\hat{\theta}, \theta) = \|\theta - \hat{\theta}\|^2$

devise an estimator $\hat{\theta} = \hat{\theta}(X)$ *such that the risk*

$$R(\hat{\theta}, \theta) := \mathbb{E}_\theta(W(\hat{\theta}, \theta)) = \int_{\mathcal{X}} W(\hat{\theta}(x), \theta)\, \mathbb{P}_\theta(dx)$$

is small.

Parametric estimation

Problem

Given

- *a subset* Θ *of* $\mathbb{R}^k$
- *data* $X \sim \mathbb{P}_\theta$ *with* $\theta \in \Theta$ *and* $\mathbb{P}_\theta$ *probability distribution on* $(\mathcal{X}, \Sigma)$
- *a loss function* $W : \Theta \times \Theta \to \mathbb{R}_+$, *e.g.* $W(\hat{\theta}, \theta) = \|\theta - \hat{\theta}\|^2$

devise an estimator $\hat{\theta} = \hat{\theta}(X)$ *such that the risk*

$$R(\hat{\theta}, \theta) := \mathbb{E}_\theta(W(\hat{\theta}, \theta)) = \int_{\mathcal{X}} W(\hat{\theta}(x), \theta)\, \mathbb{P}_\theta(dx)$$

is small.

Remark

- The same problem can be formulated for 'non-parametric' Θ, and/or estimation of a function $t = t(\theta)$
- In general the estimator may be randomised, for example
  - $\hat{\theta} = \hat{\theta}(X, U)$ where U is an additional random variable with fixed, known distribution

Unbiased estimators

Bias-variance trade-off (exercise)

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model and let $X \sim \mathbb{P}_\theta$. The mean square error of $\hat{\theta}(X)$ is the sum of a variance and a bias term

$$\begin{aligned}\mathbb{E}_\theta((\hat{\theta}-\theta)^2) &= \int (\hat{\theta}(x)-\theta)^2 \mathbb{P}_\theta(dx) = \\ &\int (\hat{\theta}(x) - \mathbb{E}_\theta(\hat{\theta}))^2 \mathbb{P}_\theta(dx) + (\theta - \mathbb{E}_\theta(\hat{\theta}))^2 = V(\hat{\theta}) + B(\hat{\theta})^2\end{aligned}$$

Definition

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model and let $X \sim \mathbb{P}_\theta$.
An estimator $\hat{\theta}(X)$ is called unbiased if $\mathbb{E}_\theta(\hat{\theta}(X)) = \theta$ for all θ.
If $\hat{\theta}$ is unbiased then the mean square error is equal to $V(\hat{\theta})$.

Bias-variance trade-off (exercise)

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model and let $X \sim \mathbb{P}_\theta$. The mean square error of $\hat{\theta}(X)$ is the sum of a variance and a bias term

$$\mathbb{E}_\theta((\hat{\theta} - \theta)^2) = \int (\hat{\theta}(x) - \theta)^2 \mathbb{P}_\theta(dx) = \int (\hat{\theta}(x) - \mathbb{E}_\theta(\hat{\theta}))^2 \mathbb{P}_\theta(dx) + (\theta - \mathbb{E}_\theta(\hat{\theta}))^2 = V(\hat{\theta}) + B(\hat{\theta})^2$$

Definition

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model and let $X \sim \mathbb{P}_\theta$. An estimator $\hat{\theta}(X)$ is called unbiased if $\mathbb{E}_\theta(\hat{\theta}(X)) = \theta$ for all θ. If $\hat{\theta}$ is unbiased then the mean square error is equal to $V(\hat{\theta})$.

Example

- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli with $\mathbb{P}_\theta([X = 1]) = \theta$ and $\mathbb{P}_\theta([X = 0]) = 1 - \theta$. Then $\bar{X} = (\sum X_i)/n$ is an unbiased estimator of θ

Unbiased estimators

Bias-variance trade-off (exercise)

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model and let $X \sim \mathbb{P}_\theta$.
The mean square error of $\hat{\theta}(X)$ is the sum of a variance and a bias term

$$\begin{aligned}\mathbb{E}_\theta((\hat{\theta}-\theta)^2) &= \int (\hat{\theta}(x)-\theta)^2 \mathbb{P}_\theta(dx) = \\ &\int (\hat{\theta}(x)-\mathbb{E}_\theta(\hat{\theta}))^2 \mathbb{P}_\theta(dx) + (\theta - \mathbb{E}_\theta(\hat{\theta}))^2 = V(\hat{\theta}) + B(\hat{\theta})^2\end{aligned}$$

Definition

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model and let $X \sim \mathbb{P}_\theta$.
An estimator $\hat{\theta}(X)$ is called unbiased if $\mathbb{E}_\theta(\hat{\theta}(X)) = \theta$ for all θ.
If $\hat{\theta}$ is unbiased then the mean square error is equal to $V(\hat{\theta})$.

Example

- Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli with $\mathbb{P}_\theta([X=1]) = \theta$ and $\mathbb{P}_\theta([X=0]) = 1-\theta$. Then $\bar{X} = (\sum X_i)/n$ is an unbiased estimator of θ
- Let $Y_1, \ldots, Y_n$ be i.i.d. normal distributed with $P_\theta = N(\theta, V)$. Then $\bar{Y} = (\sum Y_i)/n$ is an unbiased estimator of θ

Fisher information matrix

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model with $\mathbb{P}_\theta$ probability measures on $(\mathcal{X}, \Sigma)$ dominated by μ, i.e. $\mu(A) = 0 \Rightarrow \mathbb{P}_\theta(A) = 0$ for all θ.

Smooth model

Throughout we will assume that the probability densities $p_\theta = \frac{d\mathbb{P}_\theta}{d\mu}$ satisfy sufficient 'regularity conditions' allowing for differentiation w.r.t. θ and exchangeability of integral and derivative.

Fisher information matrix

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model with $\mathbb{P}_\theta$ probability measures on $(\mathcal{X}, \Sigma)$ dominated by μ, i.e. $\mu(A) = 0 \Rightarrow \mathbb{P}_\theta(A) = 0$ for all θ.

Smooth model

Throughout we will assume that the probability densities $p_\theta = \frac{d\mathbb{P}_\theta}{d\mu}$ satisfy sufficient 'regularity conditions' allowing for differentiation w.r.t. θ and exchangeability of integral and derivative.

Definition

Let $\ell_\theta := \log p_\theta$ be the log likelihood and let $\dot{\ell}_{\theta,i} := \partial \ell_\theta / \partial \theta_i$ be the score function(s).

The Fisher information matrix is defined by

$$I_{i,j}(\theta) := \mathbb{E}_\theta(\dot{\ell}_{\theta,i}\dot{\ell}_{\theta,j}) = \int_{\mathrm{supp}(p_\theta)} p_\theta^{-1}(x)\frac{\partial p_\theta}{\partial \theta_i}(x)\frac{\partial p_\theta}{\partial \theta_j}(x)\mu(dx)$$

where $\mathrm{supp}(p_\theta) = \{x : p_\theta(x) > 0\}$.

Theorem (Cramér-Rao)

Let $\hat{\theta}$ be an unbiased estimator of θ. Then the following matrix inequality holds

$$\mathbb{E}_\theta((\hat{\theta}-\theta)^T(\hat{\theta}-\theta)) = Var(\hat{\theta}) \geq I(\theta)^{-1}$$

where $I(\theta)$ is the Fisher information matrix.

The Cramer-Rao bound

Theorem (Cramér-Rao)

Let $\hat{\theta}$ be an unbiased estimator of θ. Then the following matrix inequality holds

$$\mathbb{E}_\theta((\hat{\theta}-\theta)^T(\hat{\theta}-\theta)) = Var(\hat{\theta}) \geq I(\theta)^{-1}$$

where $I(\theta)$ is the Fisher information matrix.

Proof.

Let θ be one dimensional. The general case is left as exercise.
By Cauchy-Schwarz

$$Var(\hat{\theta}) \cdot I(\theta) = \mathbb{E}_\theta((\hat{\theta}-\theta)^2) \cdot \mathbb{E}_\theta(\dot{\ell}_\theta^2) \geq \left|\mathbb{E}_\theta((\hat{\theta}-\theta)\dot{\ell}_\theta)\right|^2$$

The right side is

$$\begin{aligned} & \mathbb{E}_\theta((\hat{\theta}-\theta)\dot{\ell}_\theta) = \mathbb{E}_\theta(\hat{\theta}\dot{\ell}_\theta) - \theta\mathbb{E}_\theta(\dot{\ell}_\theta) = \\ = & \int \hat{\theta}(x)\frac{dp_\theta}{d\theta}(x)\mu(dx) - \theta\int \frac{dp_\theta}{d\theta}(x)\mu(dx) = \\ = & \frac{d}{d\theta}\int \hat{\theta}(x)p_\theta(x)\mu(dx) - \theta\frac{d}{d\theta}\int p_\theta(x)\mu(dx) = \frac{d}{d\theta}\mathbb{E}_\theta(\hat{\theta}) = 1 \quad \square \end{aligned}$$

Properties of the Fisher information matrix

- $I(\theta)$ is a positive definite real $k \times k$ matrix
- $I(\theta)$ is additive for products of independent models (exercise):
  if $\mathbb{P}_\theta = \mathbb{P}_\theta^{(1)} \times \mathbb{P}_\theta^{(2)}$ then $I(\theta) = I^{(1)}(\theta) + I^{(2)}(\theta)$

Properties of the Fisher information matrix

- $I(\theta)$ is a positive definite real $k \times k$ matrix
- $I(\theta)$ is additive for products of independent models (exercise):
  if $\mathbb{P}_\theta = \mathbb{P}_\theta^{(1)} \times \mathbb{P}_\theta^{(2)}$ then $I(\theta) = I^{(1)}(\theta) + I^{(2)}(\theta)$
- The Hellinger distance between infinitesimally close densities p_θ and $p_{\theta+d\theta}$ is determined by the Fisher information

$$h(p_\theta, p_{\theta+d\theta})^2 = \int (\sqrt{p_\theta(x)} - \sqrt{p_{\theta+d\theta}(x)})^2 \mu(dx) = \frac{1}{4} I(\theta)(d\theta)^2 + o((d\theta)^2)$$

Properties of the Fisher information matrix

- $I(\theta)$ is a positive definite real $k \times k$ matrix
- $I(\theta)$ is additive for products of independent models (exercise):
  if $\mathbb{P}_\theta = \mathbb{P}_\theta^{(1)} \times \mathbb{P}_\theta^{(2)}$ then $I(\theta) = I^{(1)}(\theta) + I^{(2)}(\theta)$
- The Hellinger distance between infinitesimally close densities p_θ and $p_{\theta+d\theta}$ is determined by the Fisher information

$$h(p_\theta, p_{\theta+d\theta})^2 = \int (\sqrt{p_\theta(x)} - \sqrt{p_{\theta+d\theta}(x)})^2 \mu(dx) = \frac{1}{4} I(\theta)(d\theta)^2 + o((d\theta)^2)$$

- The Fisher information matrix defines a riemannian metric on Θ and the corresponding geodesic distance is the Bhattacharya distance

$$d(p_{\theta_1}, p_{\theta_2}) = 2 \arccos \left(\int \sqrt{p_{\theta_1}(x)} \sqrt{p_{\theta_2}(x)} \mu(dx) \right)$$

Properties of the Fisher information matrix

- $I(\theta)$ is a positive definite real $k \times k$ matrix
- $I(\theta)$ is additive for products of independent models (exercise):
  if $\mathbb{P}_\theta = \mathbb{P}_\theta^{(1)} \times \mathbb{P}_\theta^{(2)}$ then $I(\theta) = I^{(1)}(\theta) + I^{(2)}(\theta)$
- The Hellinger distance between infinitesimally close densities p_θ and $p_{\theta+d\theta}$ is determined by the Fisher information

$$h(p_\theta, p_{\theta+d\theta})^2 = \int (\sqrt{p_\theta(x)} - \sqrt{p_{\theta+d\theta}(x)})^2 \mu(dx) = \frac{1}{4} I(\theta)(d\theta)^2 + o((d\theta)^2)$$

- The Fisher information matrix defines a riemannian metric on Θ and the corresponding geodesic distance is the Bhattacharya distance

$$d(p_{\theta_1}, p_{\theta_2}) = 2 \arccos \left(\int \sqrt{p_{\theta_1}(x)} \sqrt{p_{\theta_2}(x)} \mu(dx) \right)$$

- Let q_θ be the probability density of a randomisation Y of X (randomised statistic, Markov kernel) where $X \sim \mathbb{P}_\theta$. Then

$$d(q_{\theta_1}, q_{\theta_2}) \leq d(p_{\theta_1}, p_{\theta_2}) \qquad \text{and} \qquad h(q_{\theta_1}, q_{\theta_2}) \leq h(p_{\theta_1}, p_{\theta_2})$$

- $I(\theta)$ is the unique metric contracting under *all* randomisations

Remarks on the Cramér-Rao bound

- One can similarly define unbiased estimators $\hat{g}$ of $g(\theta)$ for a function $g : \Theta \to \mathbb{R}^p$. The Cramér-Rao bound in this case is

$$Var(\hat{g}) \geq J(\theta) I(\theta)^{-1} J(\theta)^T$$

where $J(\theta)_{l,i} = \partial g(\theta)_l / \partial \theta_i$ is the $p \times k$ Jacobian matrix (exercise).

Remarks on the Cramér-Rao bound

- One can similarly define unbiased estimators $\hat{g}$ of $g(\theta)$ for a function $g : \Theta \to \mathbb{R}^p$. The Cramér-Rao bound in this case is

$$Var(\hat{g}) \geq J(\theta) I(\theta)^{-1} J(\theta)^T$$

  where $J(\theta)_{l,i} = \partial g(\theta)_l / \partial \theta_i$ is the $p \times k$ Jacobian matrix (exercise).

- For certain models there exist no unbiased estimators, e.g. the binomial distribution $b(\theta, n)$ and function $g(\theta) = \theta^{-1}$ (exercise).

Remarks on the Cramér-Rao bound

- One can similarly define unbiased estimators $\hat{g}$ of $g(\theta)$ for a function $g : \Theta \to \mathbb{R}^p$. The Cramér-Rao bound in this case is

$$Var(\hat{g}) \geq J(\theta)I(\theta)^{-1}J(\theta)^T$$

  where $J(\theta)_{l,i} = \partial g(\theta)_l / \partial \theta_i$ is the $p \times k$ Jacobian matrix (exercise).

- For certain models there exist no unbiased estimators, e.g. the binomial distribution $b(\theta, n)$ and function $g(\theta) = \theta^{-1}$ (exercise).

- Even if unbiased estimators exist, their variance may be too big.

Remarks on the Cramér-Rao bound

- One can similarly define unbiased estimators $\hat{g}$ of $g(\theta)$ for a function $g : \Theta \to \mathbb{R}^p$. The Cramér-Rao bound in this case is

$$Var(\hat{g}) \geq J(\theta)I(\theta)^{-1}J(\theta)^T$$

  where $J(\theta)_{l,i} = \partial g(\theta)_l / \partial \theta_i$ is the $p \times k$ Jacobian matrix (exercise).

- For certain models there exist no unbiased estimators, e.g. the binomial distribution $b(\theta, n)$ and function $g(\theta) = \theta^{-1}$ (exercise).

- Even if unbiased estimators exist, their variance may be too big.

- The Cramér-Rao bound is in general not attainable, but it becomes equality if and only if the distributions form an exponential family:

  $\hat{g}$ is an unbiased estimator of $g(\theta)$ which attains CR iff

$$\frac{d \log p_\theta(x)}{d\theta} = a(\theta)(\hat{g}(x) - g(\theta))$$

Asymptotic efficiency

The theory of asymptotic efficiency shows that the Cramér-Rao bound is asymptotically attained in the following sense.

Definition

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model. Let $X_1, \ldots, X_n$ be i.i.d. with distribution $\mathbb{P}_\theta$. An estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ is called asymptotically efficient if

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, I(\theta)^{-1})$$

In particular, $\hat{\theta}_n$ attains the CR bound asymptotically:

$$n\mathbb{E}_\theta \left[(\hat{\theta}_n - \theta)^T(\hat{\theta}_n - \theta)\right] \to I(\theta)^{-1}.$$

Asymptotic efficiency

The theory of asymptotic efficiency shows that the Cramér-Rao bound is asymptotically attained in the following sense.

Definition

Let $\{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric statistical model. Let $X_1, \dots, X_n$ be i.i.d. with distribution $\mathbb{P}_\theta$. An estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ is called asymptotically efficient if

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, I(\theta)^{-1})$$

In particular, $\hat{\theta}_n$ attains the CR bound asymptotically:

$$n\mathbb{E}_\theta \left[(\hat{\theta}_n - \theta)^T (\hat{\theta}_n - \theta) \right] \to I(\theta)^{-1}.$$

Theorem

Under regularity conditions, the maximum likelihood estimator

$$\hat{\theta}_n(X_1, \dots, X_n) = \arg\max_\tau \prod_{i=1}^n p_\tau(X_i) = \arg\max_\tau \sum_{i=1}^n \ell_\tau(X_i)$$

is asymptotically efficient.

Maximum likelihood is asymptotically efficient: ideas of the proof

Log likelihood: $\ell_n(\theta) = \ell_{\theta,n}(X_1, \dots X_n) = \sum_{i=1}^n \ell_\theta(X_i)$

1. Let θ_0 be the true parameter. If $\ell_n(\theta)$ is twice differentiable around θ_0 then
   - Consistency: $\hat{\theta}_n \to \theta_0$ with probability one
   - $\dot{\ell}_n(\hat{\theta}_n) = 0$ since $\hat{\theta}_n$ is maximum likelihood

2. By Taylor expansion there is a θ_n^* between θ_0 and $\hat{\theta}_n$ such that

$$-\dot{\ell}_n(\theta_0) = \dot{\ell}_n(\hat{\theta}_n) - \dot{\ell}_n(\theta_0) = (\hat{\theta}_n - \theta_0)\ddot{\ell}_n(\theta_n^*)$$

from which

$$\hat{\theta}_n - \theta_0 = -\frac{\dot{\ell}_n(\theta_0)}{\ddot{\ell}_n(\theta_n^*)}$$

so

$$\sqrt{nI(\theta_0)}\left(\hat{\theta}_n - \theta_0\right) = \frac{\dot{\ell}_n(\theta_0)}{\sqrt{nI(\theta_0)}} \cdot \frac{\ddot{\ell}_n(\theta_0)}{\ddot{\ell}_n(\theta_n^*)} \cdot \left(-\frac{\ddot{\ell}_n(\theta_0)}{nI(\theta_0)}\right)^{-1}$$

3. The right side converges in distribution to $N(0,1)$ since
   - by C. L. T.: $\dot{\ell}_n(\theta_0)/\sqrt{nI(\theta_0)} \xrightarrow{\mathcal{L}} N(0,1)$ using $\mathbb{E}_{\theta_0}(\dot{\ell}^2_{\theta_0}) = I(\theta_0)$
   - by L. L. N. the third term converges to 1 since $\mathbb{E}_{\theta_0}(\ddot{\ell}_{\theta_0}) = -I(\theta_0)$
   - the middle term converges to one by using consistency of $\hat{\theta}_n$

Exercise: Maximum likelihood for Gaussian models

Let $\mathbb{P}_{(x,y)} := N((x,y),V)$ be Gaussian model with unknown mean $(x,y) \in \mathbb{R}^2$ and known, (non-degenerate) covariace matrix V.

1. Let $(X,Y) \sim \mathbb{P}_{(x,y)}$. Show that the Fisher information is $I = V^{-1}$ and the maximum likelihood estimator of (x,y) is $(\hat{x},\hat{y}) = (X,Y)$, and achieves the Cramer-Rao bound. In particular

$$\mathbb{E}(\hat{x}-x)^2 = V_{11} = (I^{-1})_{11}$$

2. Consider that y is known, e.g. $y=0$ and we would like to estimate x from $(X,Y) \sim \mathbb{P}_{(x,0)}$. Find the maximum likelihood estimator $\tilde{x}$ and show that

$$\mathbb{E}(\tilde{x}-x)^2 = (I_{11})^{-1} \leq (I^{-1})_{11}$$

Example: linear regression and least squares

Problem (Linear regression)

estimate the unknown vector $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ given observations

$$Y_i = \sum_j A_{ij} x_j + \epsilon_i$$

with known A_{ij} and i.i.d $\epsilon_i \sim N(0, \sigma^2)$.

Example: linear regression and least squares

Problem (Linear regression)
estimate the unknown vector $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ given observations

$$Y_i = \sum_j A_{ij} x_j + \epsilon_i$$

with known A_{ij} and i.i.d $\epsilon_i \sim N(0, \sigma^2)$.

Least squares: Find $\hat{x}$ which minimises

$$\sum_i |Y_i - \sum_j A_{ij} \hat{x}_j|^2 = (\mathbf{Y} - A\hat{\mathbf{X}})^T (\mathbf{Y} - A\hat{\mathbf{X}})$$

Example: linear regression and least squares

Problem (Linear regression)
estimate the unknown vector $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ given observations

$$Y_i = \sum_j A_{ij} x_j + \epsilon_i$$

with known A_{ij} and i.i.d $\epsilon_i \sim N(0, \sigma^2)$.

Least squares: Find $\hat{x}$ which minimises

$$\sum_i |Y_i - \sum_j A_{ij} \hat{x}_j|^2 = (\mathbf{Y} - A\hat{\mathbf{X}})^T (\mathbf{Y} - A\hat{\mathbf{X}})$$

Explicit solution coinciding with maximum likelihood estimator

$$\hat{\mathbf{X}} = (A^T A)^{-1} A^T \mathbf{Y}$$

Covariance matrix of $\hat{X}$

$$Var(\hat{\mathbf{X}}) = \sigma^2 (A^T A)^{-1}$$

Example: repeated coin toss

Let $\mathbb{P}_\theta$ be the Bernoulli distribution: $\mathbb{P}_\theta([X = 1]) = \theta$ and $\mathbb{P}_\theta([X = 0]) = 1 - \theta$.
Let $X_1, \ldots, X_n$ be i.i.d. with distribution $\mathbb{P}_\theta$. Then

Example: repeated coin toss

Let $\mathbb{P}_\theta$ be the Bernoulli distribution: $\mathbb{P}_\theta([X = 1]) = \theta$ and $\mathbb{P}_\theta([X = 0]) = 1 - \theta$. Let $X_1, \ldots, X_n$ be i.i.d. with distribution $\mathbb{P}_\theta$. Then

- $\bar{X}_n := (\sum_{i=1}^n X_i)/n$ is an unbiased estimator of θ. Indeed

$$\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}(X) = \mathbb{P}_\theta([X = 0]) \cdot 0 + \mathbb{P}_\theta([X = 1]) \cdot 1 = \theta$$

Example: repeated coin toss

Let $\mathbb{P}_\theta$ be the Bernoulli distribution: $\mathbb{P}_\theta([X=1]) = \theta$ and $\mathbb{P}_\theta([X=0]) = 1-\theta$. Let $X_1, \ldots, X_n$ be i.i.d. with distribution $\mathbb{P}_\theta$. Then

- $\bar{X}_n := (\sum_{i=1}^n X_i)/n$ is an unbiased estimator of θ. Indeed

$$\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}(X) = \mathbb{P}_\theta([X=0]) \cdot 0 + \mathbb{P}_\theta([X=1]) \cdot 1 = \theta$$

- The variance of $\bar{X}_n$ is $Var(\bar{X}_n) = Var(X)/n$ and

$$\begin{aligned} Var(X) &= \mathbb{P}_\theta([X=0]) \cdot (0 - \mathbb{E}_\theta(X))^2 + \mathbb{P}_\theta([X=1]) \cdot (1 - \mathbb{E}_\theta(X))^2 \\ &= \theta(1-\theta)^2 + (1-\theta)(0-\theta)^2 = \theta(1-\theta) \end{aligned}$$

Example: repeated coin toss

Let $\mathbb{P}_\theta$ be the Bernoulli distribution: $\mathbb{P}_\theta([X = 1]) = \theta$ and $\mathbb{P}_\theta([X = 0]) = 1 - \theta$. Let $X_1, \ldots, X_n$ be i.i.d. with distribution $\mathbb{P}_\theta$. Then

- $\bar{X}_n := (\sum_{i=1}^n X_i)/n$ is an unbiased estimator of θ. Indeed

$$\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}(X) = \mathbb{P}_\theta([X = 0]) \cdot 0 + \mathbb{P}_\theta([X = 1]) \cdot 1 = \theta$$

- The variance of $\bar{X}_n$ is $Var(\bar{X}_n) = Var(X)/n$ and

$$\begin{aligned} Var(X) &= \mathbb{P}_\theta([X = 0]) \cdot (0 - \mathbb{E}_\theta(X))^2 + \mathbb{P}_\theta([X = 1]) \cdot (1 - \mathbb{E}_\theta(X))^2 \\ &= \theta(1-\theta)^2 + (1-\theta)(0-\theta)^2 = \theta(1-\theta) \end{aligned}$$

- The Fisher information is

$$I(\theta) = \mathbb{E}_\theta[\dot{\ell}_\theta^2] = \theta^{-1} + (1-\theta)^{-1} = 1/(\theta(1-\theta))$$

Example: repeated coin toss

Let $\mathbb{P}_\theta$ be the Bernoulli distribution: $\mathbb{P}_\theta([X=1]) = \theta$ and $\mathbb{P}_\theta([X=0]) = 1-\theta$. Let $X_1, \ldots, X_n$ be i.i.d. with distribution $\mathbb{P}_\theta$. Then

- $\bar{X}_n := (\sum_{i=1}^n X_i)/n$ is an unbiased estimator of θ. Indeed

$$\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}(X) = \mathbb{P}_\theta([X=0]) \cdot 0 + \mathbb{P}_\theta([X=1]) \cdot 1 = \theta$$

- The variance of $\bar{X}_n$ is $Var(\bar{X}_n) = Var(X)/n$ and

$$\begin{aligned} Var(X) &= \mathbb{P}_\theta([X=0]) \cdot (0 - \mathbb{E}_\theta(X))^2 + \mathbb{P}_\theta([X=1]) \cdot (1 - \mathbb{E}_\theta(X))^2 \\ &= \theta(1-\theta)^2 + (1-\theta)(0-\theta)^2 = \theta(1-\theta) \end{aligned}$$

- The Fisher information is

$$I(\theta) = \mathbb{E}_\theta[\dot{\ell}_\theta^2] = \theta^{-1} + (1-\theta)^{-1} = 1/(\theta(1-\theta))$$

Thus $\bar{X}_n$ attains the Cramér-Rao bound.

Example: repeated coin toss

By the Central Limit Theorem we have

$$\sqrt{n}(\bar{X}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \theta) \xrightarrow{\mathcal{L}} N(0, Var(X)) = N(0, \theta(1-\theta))$$

Thus $\bar{X}_n$ is asymptotically efficient.

Example: repeated coin toss

By the Central Limit Theorem we have

$$\sqrt{n}(\bar{X}_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \theta) \xrightarrow{\mathcal{L}} N(0, Var(X)) = N(0, \theta(1-\theta))$$

Thus $\bar{X}_n$ is asymptotically efficient.

The maximum likelihood estimator is $\hat{\theta}_n = \bar{X}_n$! Indeed

$$\frac{dp_\theta}{d\theta}(X_1, \ldots, X_n) = \frac{d}{d\theta} \prod_{i=1}^{n} \theta^{\sum_i X_i} (1-\theta)^{n-\sum_i X_i} = \left(\frac{\sum_i X_i}{\theta} - \frac{n - \sum_i X_i}{1-\theta} \right) p_\theta = 0$$

has solution $\hat{\theta}_n = \bar{X}_n$.

Local asymptotic normality for coin toss

The random variable $\bar{X}_n \in \{0, 1/n, \ldots, n/n\}$ has binomial distribution $\mathrm{Bin}(n, \theta)$

$$\mathbb{P}_\theta[\bar{X}_n = k/n] = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

The CLT says that the (centred and rescaled) binomial is approximated by the normal $N(0, \theta(1-\theta))$ with variance depending on θ.

A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters θ of size $n^{-1/2}$

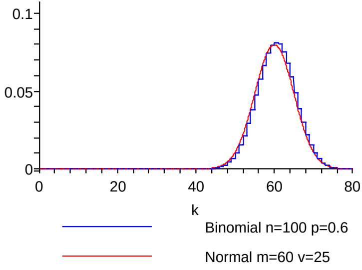
Local asymptotic normality for coin toss

The random variable $\bar{X}_n \in \{0, 1/n, \ldots, n/n\}$ has binomial distribution $\mathrm{Bin}(n, \theta)$

$$\mathbb{P}_\theta[\bar{X}_n = k/n] = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

The CLT says that the (centred and rescaled) binomial is approximated by the normal $N(0, \theta(1-\theta))$ with variance depending on θ.

A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters θ of size $n^{-1/2}$



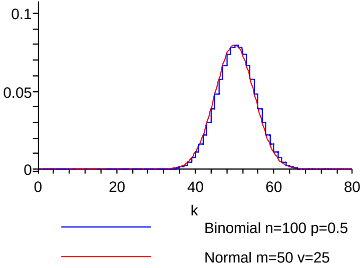
Local asymptotic normality for coin toss

The random variable $\bar{X}_n \in \{0, 1/n, \ldots, n/n\}$ has binomial distribution $\text{Bin}(n, \theta)$

$$\mathbb{P}_\theta[\bar{X}_n = k/n] = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

The CLT says that the (centred and rescaled) binomial is approximated by the normal $N(0, \theta(1-\theta))$ with variance depending on θ.

A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters θ of size $n^{-1/2}$



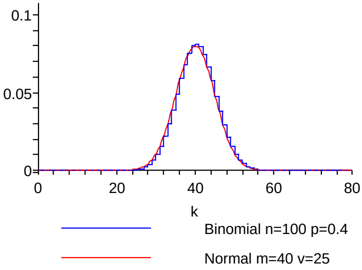
Local asymptotic normality for coin toss

The random variable $\bar{X}_n \in \{0, 1/n, \ldots, n/n\}$ has binomial distribution $\mathrm{Bin}(n, \theta)$

$$\mathbb{P}_\theta[\bar{X}_n = k/n] = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

The CLT says that the (centred and rescaled) binomial is approximated by the normal $N(0, \theta(1-\theta))$ with variance depending on θ.

A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters θ of size $n^{-1/2}$



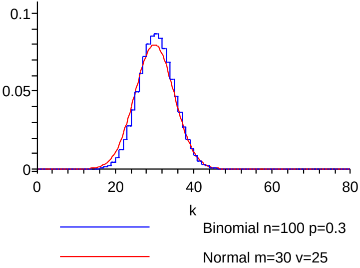
Local asymptotic normality for coin toss

The random variable $\bar{X}_n \in \{0, 1/n, \ldots, n/n\}$ has binomial distribution $\text{Bin}(n, \theta)$

$$\mathbb{P}_\theta[\bar{X}_n = k/n] = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

The CLT says that the (centred and rescaled) binomial is approximated by the normal $N(0, \theta(1-\theta))$ with variance depending on θ.

A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters θ of size $n^{-1/2}$



Local asymptotic normality for coin toss

Consider θ in a $n^{-1/2}$ neighbourhood of a fixed point θ_0 such that $\theta = \theta_0 + u/\sqrt{n}$ for some local parameter u.

Let $\hat{u}_n$ be unbiased estimator of u obtained by centering and rescaling $\bar{X}_n$

$$\hat{u}_n := \sqrt{n}(\bar{X}_n - \theta_0)$$

Lemma

For any local parameter u the convergence in distribution holds

$$\hat{u}_n \xrightarrow{\mathcal{L}} N(u, \theta_0(1-\theta_0))$$

Consider θ in a $n^{-1/2}$ neighbourhood of a fixed point θ_0 such that $\theta = \theta_0 + u/\sqrt{n}$ for some local parameter u.

Let $\hat{u}_n$ be unbiased estimator of u obtained by centering and rescaling $\bar{X}_n$

$$\hat{u}_n := \sqrt{n}(\bar{X}_n - \theta_0)$$

Lemma

For any local parameter u the convergence in distribution holds

$$\hat{u}_n \xrightarrow{\mathcal{L}} N(u, \theta_0(1-\theta_0))$$

Proof: (exercise)

Hint: by Lévy's Theorem it suffices to prove convergence of characteristic functions:

$$\mathbb{E}_{\theta_0+u/\sqrt{n}}(\exp(it\hat{u}_n)) \to \exp(itu)\cdot\exp(-t^2\theta_0(1-\theta_0)/2).$$

Since $\hat{u}_n$ is a sum of i.i.d. variables the left side is

$$\left[\mathbb{E}_{\theta_0+u/\sqrt{n}}(\exp(it(X-\theta_0)/\sqrt{n}))\right]^n = \left(1 - \frac{\theta_0(1-\theta_0)t^2/2 + itu}{n} + o(n^{-3/2})\right)^n.$$

Local asymptotic normality for coin toss

Summarising the previous two slides, we showed that asymptotically with n, the estimator $\hat{u}_n$ of the local parameter u converges in distribution to a normal with mean u and fixed variance $I(\theta_0)^{-1}$.

Local asymptotic normality for coin toss

Summarising the previous two slides, we showed that asymptotically with n, the estimator $\hat{u}_n$ of the local parameter u converges in distribution to a normal with mean u and fixed variance $I(\theta_0)^{-1}$.

Since $\hat{u}_n$ is a sufficient statistic for the data $(X_1, \ldots, X_n)$ this means that the original i.i.d. model converges (locally) to a simple Gaussian shift model

$$\{\mathbb{P}^n_{\theta_0+u/\sqrt{n}} : u \in \mathbb{R}\} \longrightarrow \{N(u, I(\theta_0)^{-1}) : u \in \mathbb{R}\}.$$

Local asymptotic normality for coin toss

Summarising the previous two slides, we showed that asymptotically with n, the estimator $\hat{u}_n$ of the local parameter u converges in distribution to a normal with mean u and fixed variance $I(\theta_0)^{-1}$.

Since $\hat{u}_n$ is a sufficient statistic for the data $(X_1, \ldots, X_n)$ this means that the original i.i.d. model converges (locally) to a simple Gaussian shift model

$$\{\mathbb{P}^n_{\theta_0+u/\sqrt{n}} : u \in \mathbb{R}\} \longrightarrow \{N(u, I(\theta_0)^{-1}) : u \in \mathbb{R}\}.$$

This phenomenon is called local asymptotic normality and holds (with an appropriate definition of convergence) for arbitrary 'smooth' statistical models $\{\mathbb{P}_\theta : \theta \in \mathbb{R}^k\}$. The theory of convergence of statistical models is a classical topic in asymptotic statistics, which can be used to find asymptotically optimal estimators and estimation rates, by transforming complicated models into simpler Gaussian ones.

Local asymptotic normality for coin toss

Summarising the previous two slides, we showed that asymptotically with n, the estimator $\hat{u}_n$ of the local parameter u converges in distribution to a normal with mean u and fixed variance $I(\theta_0)^{-1}$.

Since $\hat{u}_n$ is a sufficient statistic for the data $(X_1, \ldots, X_n)$ this means that the original i.i.d. model converges (locally) to a simple Gaussian shift model

$$\{\mathbb{P}^n_{\theta_0 + u/\sqrt{n}} : u \in \mathbb{R}\} \longrightarrow \{N(u, I(\theta_0)^{-1}) : u \in \mathbb{R}\}.$$

This phenomenon is called local asymptotic normality and holds (with an appropriate definition of convergence) for arbitrary 'smooth' statistical models $\{\mathbb{P}_\theta : \theta \in \mathbb{R}^k\}$. The theory of convergence of statistical models is a classical topic in asymptotic statistics, which can be used to find asymptotically optimal estimators and estimation rates, by transforming complicated models into simpler Gaussian ones.

Later on we will use this idea as our guiding principle in finding asymptotically optimal procedures for quantum state estimation.

Definition

Let $X \sim \mathbb{P}_\theta$ with $\theta \in \Theta \subset \mathbb{R}$.

An interval $\mathcal{I}(X) = [L(X), U(X)]$ is called confidence interval of level $1-\alpha$ if

$$\mathbb{P}_\theta(L(X) \leq \theta \leq U(X)) = 1-\alpha, \qquad \forall\theta$$

Remark

- Similar definition holds for confidence intervals for χ when $\theta = (\chi, \psi) \in \mathbb{R} \times \mathbb{R}^{k-1}$
- There exists a general procedure for constructing confidence intervals from tests for hypotheses of the type $H_0 = \{\theta = \theta_0\}$ and $H_1 = \{\theta \neq \theta_0\}$, and vice-versa
- In general it is difficult to construct exact confidence intervals and approximate intervals are used instead $\mathbb{P}_\theta(L(X) \leq \theta \leq U(X)) \approx 1-\alpha$

Confidence intervals (exercise)

Given $X_1, \ldots X_n$ i.i.d. with $N(\mu, \sigma^2)$, show that

- the sample mean $\overline{X}_n = \sum X_i/n$ satisfies

$$\sqrt{n}(\overline{X}_n - \mu) \sim N(0, \sigma^2)$$

- the sample variance $s_n^2 = \sum (X_i - \overline{X}_n)^2/(n-1)$ satisfies

$$(n-1)s_n^2/\sigma^2 \sim \chi^2_{n-1}$$

  where χ^2_{n-1} is the chi-square with (n-1) degrees of freedom, i.e. the distribution of $\sum_{j=1}^{n-1} Y_j^2$ with $Y_i \sim N(0,1)$ independent

- $\overline{X}_n$ and s_n^2 are independent

- from the above follows that

$$T_n := \frac{\sqrt{n}(\overline{X}_n - \mu)}{s_n} \sim t_{n-1} \qquad (*)$$

  with t_{n-1} denoting the student-t distribution.

- If c is taken such that $\mathbb{P}(|T_n| > c) = \alpha$ then $(*)$ implies

$$\mathbb{P}_{\mu,\sigma}\left(\bar{X}_n - \frac{cs_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n - \frac{cs_n}{\sqrt{n}}\right) = 1 - \alpha$$

  which provides a level α confidence interval for μ.

(Approximate) confidence intervals from asymptotic efficiency

Let $X_1, \ldots, X_n$ be independent with $X_i \sim \mathbb{P}_\theta$ and $\theta \in \Theta \subset \mathbb{R}$.

Recall that the maximum likelihood estimator $\hat{\theta}_n$ is asymptotically efficient

$$\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0,1)$$

From this we get that if c is such that $\mathbb{P}(|Y| > c) = \alpha$ for $Y \sim N(0,1)$ then

$$\mathbb{P}_\theta\left(\hat{\theta}_n - \frac{c}{\sqrt{nI(\hat{\theta}_n)}} \leq \theta \leq \hat{\theta}_n + \frac{c}{\sqrt{nI(\hat{\theta}_n)}}\right) \approx 1 - \alpha.$$

Remark

There are arguments for replacing the Fisher information $nI(\hat{\theta}_n)$ with the observed Fisher information $j_n(\hat{\theta}_n)$ where

$$j_n(\theta) := \sum_{i=1}^{n} \ddot{\ell}_n(\theta) = \sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log p_\theta(X_i)$$

Exercise: If $X_i \sim \mathrm{Bernoulli}(\theta)$, show that

$$I_n(\hat{\theta}_n) = j_n(\hat{\theta}_n) = \frac{n}{\hat{\theta}_n(1 - \hat{\theta}_n)}$$ $\Leftarrow$ problematic for $p(1-p) \approx 0$!

Confidence intervals by bootstrap

Bootstrap methods can be used to (approximately) sample from the distribution of an estimator or compute confidence intervals.

In parametric bootstrap we assume $X_i \sim \mathbb{P}_\theta$ with $\theta \in \Theta \subset \mathbb{R}^k$ as opposed to arbitrary distribution. The general procedure has the following steps:

1. Construct maximum likelihood estimator $\hat{\theta}_n$ from the data $X_1, \ldots, X_n$
2. Generate new i.i.d. datasets $\tilde{\mathbf{X}}^{(j)} = (\tilde{X}_1^{(j)}, \ldots, \tilde{X}_n^{(j)})$ with $j = 1, \ldots m$ and
$$\tilde{X}_i^{(j)} \sim \mathbb{P}_{\hat{\theta}_n}, \qquad \forall i, j$$
3. Compute the ml estimator $\tilde{\theta}_n^{(j)}$ for each dataset $\tilde{\mathbf{X}}^{(j)}$
4. Construct confidence intervals from the empirical distribution of the ml estimators.

Hypothesis testing

Problem

Let $\{\mathbb{P}_0, \mathbb{P}_1\}$ *be a binary statistical model over* $(\mathcal{X}, \Sigma)$. *Given* $X \sim \mathbb{P}_i$ *decide which of the two hypotheses is true,* $\mathbb{P}_0$ *or* $\mathbb{P}_1$. *A test is a function* $t : \Omega \to \{0, 1\}$ *and it's 'goodness' is measured in terms of the error probabilities*

- *type I error* $\mathbb{P}_0([t(X) = 1])$
- *type II error* $\mathbb{P}_1([t(X) = 0])$

Hypothesis testing

Problem

Let $\{\mathbb{P}_0, \mathbb{P}_1\}$ be a binary statistical model over $(\mathcal{X}, \Sigma)$. Given $X \sim \mathbb{P}_i$ decide which of the two hypotheses is true, $\mathbb{P}_0$ or $\mathbb{P}_1$. A test is a function $t : \Omega \to \{0, 1\}$ and it's 'goodness' is measured in terms of the error probabilities

- *type I error* $\mathbb{P}_0([t(X) = 1])$
- *type II error* $\mathbb{P}_1([t(X) = 0])$

Let p_0 and p_1 be the densities of $\mathbb{P}_0$ and $\mathbb{P}_1$ with respect to a measure μ.

- The likelihood ratio statistic $LR : \Omega \to \mathbb{R}$ is defined as

$$LR(x) = \frac{p_1(x)}{p_0(x)}$$

  and is a sufficient statistic for the model $\{\mathbb{P}_0, \mathbb{P}_1\}$

  Exercise: prove this for two binomials $\mathrm{Bin}(n, \theta_0)$ and $\mathrm{Bin}(n, \theta_1)$.

- The likelihood ratio test t_k is defined by

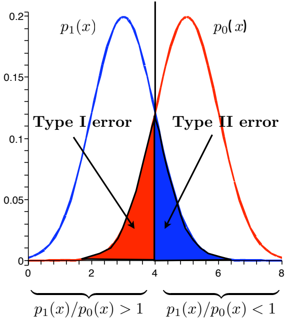
$$t_k(\omega) := \begin{cases} 0 & \text{if } p_0(x)/p_1(\omega) > k \\ 1 & \text{if } p_0(x)/p_1(\omega) \leq k \end{cases}$$

Hypothesis testing

Problem

Let $\{\mathbb{P}_0, \mathbb{P}_1\}$ be a binary statistical model over $(\mathcal{X}, \Sigma)$. Given $X \sim \mathbb{P}_i$ decide which of the two hypotheses is true, $\mathbb{P}_0$ or $\mathbb{P}_1$. A test is a function $t : \Omega \to \{0, 1\}$ and it's 'goodness' is measured in terms of the error probabilities

- *type I error* $\mathbb{P}_0([t(X) = 1])$
- *type II error* $\mathbb{P}_1([t(X) = 0])$



The likelihood ratio test t_1 for two Gaussian distributions

Optimal tests

Let $\{\mathbb{P}_0, \mathbb{P}_1\}$ be a binary statistical model over $(\mathcal{X}, \Sigma)$ and let p_0 and p_1 be the densities of $\mathbb{P}_0$ and $\mathbb{P}_1$ w.r.t. a probability measure μ.

Optimal tests

Let $\{\mathbb{P}_0, \mathbb{P}_1\}$ be a binary statistical model over $(\mathcal{X}, \Sigma)$ and let p_0 and p_1 be the densities of $\mathbb{P}_0$ and $\mathbb{P}_1$ w.r.t. a probability measure μ.

Lemma (Neyman-Pearson lemma)

Let $\alpha \in (0,1)$ be a fixed level. Then there exist a constant k such that the likelihood ratio test t_k is of level α (i.e. $\mathbb{P}_0([t(X) = 1]) = \alpha$) and minimises the type II error $\mathbb{P}_1([t(X) = 0])$ among the α-level tests.

Let $\{\mathbb{P}_0, \mathbb{P}_1\}$ be a binary statistical model over $(\mathcal{X}, \Sigma)$ and let p_0 and p_1 be the densities of $\mathbb{P}_0$ and $\mathbb{P}_1$ w.r.t. a probability measure μ.

Lemma (Neyman-Pearson lemma)

Let $\alpha \in (0,1)$ be a fixed level. Then there exist a constant k such that the likelihood ratio test t_k is of level α (i.e. $\mathbb{P}_0([t(X) = 1]) = \alpha$) and minimises the type II error $\mathbb{P}_1([t(X) = 0])$ among the α-level tests.

Lemma (optimal Bayes test)

Let π_0, π_1 be a (non-degenerate) prior distribution. Then the likelihood ratio test

$$t(\omega) := \begin{cases} 0 & \text{if } p_0(\omega)/p_1(\omega) > \pi_1/\pi_0 \\ 1 & \text{if } p_0(\omega)/p_1(\omega) \leq \pi_1/\pi_0 \end{cases}$$

has minimal average error

$$P_\pi^e := \pi_0 \mathbb{P}_0([t(X) = 1]) + \pi_1 \mathbb{P}_1([t(X) = 0]) = \frac{1}{2}(1 - \|\pi_1 p_1 - \pi_0 p_0\|_1)$$

Let $\{\mathbb{P}_0, \mathbb{P}_1\}$ be a binary statistical model and let $X_1, \ldots, X_n$ i.i.d. with $X_k \sim \mathbb{P}_i$.

Theorem (Stein's Lemma)

Let $t_n(X_1, \ldots, X_n)$ be the most powerful level α test. Then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_1^n([t_n = 0]) = -D(p_0, p_1)$$

where $D(p_0, p_1)$ is the relative entropy

$$D(p_0, p_1) = \int p_0(\omega) \log(p_0/p_1)\mu(d\omega).$$

Theorem (Chernoff's bound)

Let π_0, π_1 be a nondegenerate prior and let $t_n(X_1, \ldots, X_n)$ be the optimal Bayes test. Then

$$\lim_{n\to\infty} \frac{1}{n} \log P_\pi^{e,n} = -C(p_0, p_1)$$

where $C(p_0, p_1)$ is the Chernoff distance

$$C(p_0, p_1) = -\log\left(\inf_{0\le s\le 1} \int p_0^s(\omega) p_1^{1-s}(\omega)\mu(d\omega)\right)$$

Wilk's test

Let $X_1, \ldots, X_n \sim \mathbb{P}_\theta$ with $\theta \in \mathbb{R}^{k+m}$.
We would like to test between the two hypotheses

- $H_0 : \theta \in \Theta_0 := \{\theta : \theta_i = \theta_i^{(0)}, i = 1, \ldots, m\}$
- $H_1 : \theta \notin \Theta_0$

Let ℓ_0 and ℓ_1 be the maximum log-likelihoods over Θ_0 and Θ

$$\ell_0 := \sup\{\ell_\theta : \theta \in \Theta_0\} \qquad \ell_1 := \sup\{\ell_\theta : \theta \in \Theta\}$$

Wilk's Theorem Suppose H_0 is true. Then under regularity assumptions, the likelihood ratio statistic $T_n := 2(\ell_1 - \ell_0)$ has asymptotic χ^2_m distribution:

$$T_n \xrightarrow{\mathcal{L}} \chi^2_m$$

The theorem suggests the following test of (asymptotic) level α:

accept H_0 if $T_n > c$ where $\mathbb{P}(Y > c) = \alpha$ for $Y \sim \chi^2_m$

2. Basics of quantum state estimation

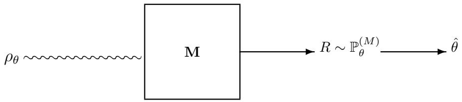
- Quantum statistical models and state estimation
- Estimation of qubit states - simple non-adaptive measurements 3D
- Estimation of qubit states - simple non-adaptive measurements 1D
- Estimation of qubit states - simple adaptive measurements 2D
- Estimation of qubit states - simple adaptive measurements 3D

Set-up of quantum estimation problems

Quantum statistical model over Θ:

$$\mathcal{Q} = \left\{ \rho^{\theta} \ : \ \theta \in \Theta \right\}$$

Estimation procedure: measure state ρ^{θ} and devise estimator $\hat{\theta} = \hat{\theta}(R)$



Risk: $R(\hat{\theta}, \theta) = \mathbb{E}_\theta[W(\hat{\theta}, \theta)]$, e.g. $W(\hat{\theta}, \theta) = \|\hat{\theta} - \theta\|^2$

Measurement design:

- which classical model $\mathcal{P}^{(M)} = \{\mathbb{P}_\theta^{(M)} : \theta \in \Theta\}$ is 'best' ?
- trade-off between incompatible observables
- optimal measurement depends on statistical problem

Quantum statistical models

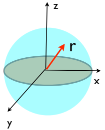
Definition

Let Θ be a parameter space. A quantum statistical model over Θ is a family $\{\rho_\theta : \theta \in \Theta\}$ of density matrices on a given space $\mathcal{H}$.

Example

- qubit states: indexed by $\mathbf{r} = (r_x, r_y, r_z) \in \mathbb{R}^3$ such that $\|\mathbf{r}\| \leq 1$

$$\rho_{\mathbf{r}} = \frac{1}{2}\begin{pmatrix} 1 + r_z & r_x - ir_y \\ r_x + ir_y & 1 - r_z \end{pmatrix}$$



- coherent spin states: $\rho_{\mathbf{r}}^n = \rho_{\mathbf{r}} \otimes \cdots \otimes \rho_{\mathbf{r}}$, for $\|\mathbf{r}\| = 1$ (pure states)
- Unitary family: $\rho_t = \exp(iHt)\rho\exp(iHt)$ for $t \in \mathbb{R}$, H selfadjoint
- Gaussian states $\Phi(z, V)$ of a quantum continuous variables system, with mean $z \in \mathbb{C}$, and 2×2 covariance matrix V

Problem

Given

- *a quantum statistical model* $\{\rho_\theta : \theta \in \Theta\}$
- *a loss function* $W : \Theta \times \Theta \to \mathbb{R}_+$, *e.g.* $\|\hat{\theta} - \theta\|^2$ *for* $\Theta \subset \mathbb{R}^k$ *or* $\|\hat{\rho} - \rho\|_1$ *if* $\Theta \subset \mathcal{S}(\mathcal{H})$, *etc.*

design a measurement M *and an estimator* $\hat{\theta}(X)$, *where* X *is the outcome of the measurement, such that*

$$R(M, \hat{\theta}, \theta) = \mathbb{E}_\theta(W(\hat{\theta}(X), \theta))$$

is small.

Remark

- same problem can be formulated for estimating a function $g(\theta)$
- the main quantum feature is the optimisation over measurements step
- measurement and estimator can be 'bundled' into a measurement with values in Θ (exercise)

Example: estimation of a spin state

Problem

Estimate $\vec{r}$, given n quantum spins prepared in state

$$\rho_{\vec{r}} := \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix} = \frac{1}{2}\left(\mathbf{1} + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z\right)$$

Example: estimation of a spin state

Problem

Estimate $\vec{r}$, given n quantum spins prepared in state

$$\rho_{\vec{r}} := \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix} = \frac{1}{2}\left(\mathbf{1} + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z\right)$$

Basic solution: measure each $\sigma_{x,y,z}$ separately on $n/3$ spins

Probability distribution for σ_i measurement: $\mathbb{P}_{\rho_{\vec{r}}}[\sigma_i = \pm 1] = (1 \pm r_i)/2$

Example: estimation of a spin state

Problem

Estimate $\vec{r}$, given n quantum spins prepared in state

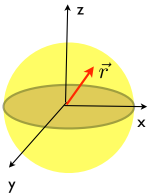
$$\rho_{\vec{r}} := \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix} = \frac{1}{2}\left(\mathbf{1} + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z\right)$$

Basic solution: measure each $\sigma_{x,y,z}$ separately on $n/3$ spins

Probability distribution for σ_i measurement: $\mathbb{P}_{\rho_{\vec{r}}}[\sigma_i = \pm 1] = (1 \pm r_i)/2$

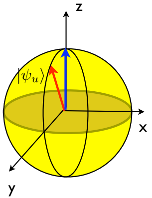
Estimator (as for coin toss) $\hat{r}_i := \frac{3}{n}(n_i^+ - n_i^-)$

Mean square error (risk) achieves the CR bound for this measurement



$$\mathbb{E}[\|\vec{r} - \hat{\vec{r}}\|^2] = \frac{3}{n}\mathrm{Tr}\left(\left(I^{(x)} + I^{(y)} + I^{(z)}\right)^{-1}\right) = \frac{3}{n}(3 - r^2)$$

Example: estimation of a pure spin state

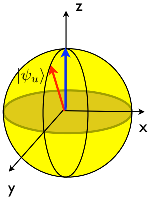


Problem

Estimate rotation parameter for the 1D model

$$|\psi_u\rangle := \exp\left(\frac{iu\sigma_x}{2}\right)|\uparrow\rangle = \cos\left(\frac{u}{2}\right)|\uparrow\rangle + i\sin\left(\frac{u}{2}\right)|\downarrow\rangle$$

Example: estimation of a pure spin state



Problem

Estimate rotation parameter for the 1D model

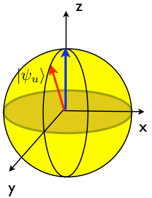
$$|\psi_u\rangle := \exp\left(\frac{iu\sigma_x}{2}\right)|\uparrow\rangle = \cos\left(\frac{u}{2}\right)|\uparrow\rangle + i\sin\left(\frac{u}{2}\right)|\downarrow\rangle$$

Measure spin observable σ_y with probabilities $\mathbb{P}[X = \pm 1] = p_u(\pm 1) = \frac{1 \pm \sin(u)}{2}$

Fisher information

$$\begin{aligned} I(u) &= \frac{1}{p_u(1)}\left(\frac{dp_u(1)}{du}\right)^2 + \frac{1}{p_u(-1)}\left(\frac{dp_u(-1)}{du}\right)^2 \\ &= \frac{\cos(u)^2}{2}\left(\frac{1}{1+\sin(u)} + \frac{1}{1-\sin(u)}\right) = 1 \end{aligned}$$

Example: estimation of a pure spin state



Problem

Estimate rotation parameter for the 1D model

$$|\psi_u\rangle := \exp\left(\frac{iu\sigma_x}{2}\right)|\uparrow\rangle = \cos\left(\frac{u}{2}\right)|\uparrow\rangle + i\sin\left(\frac{u}{2}\right)|\downarrow\rangle$$

Measure spin observable σ_y with probabilities $\mathbb{P}[X = \pm 1] = p_u(\pm 1) = \frac{1 \pm \sin(u)}{2}$

Fisher information

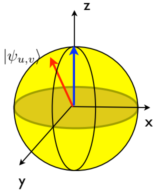
$$I(u) = \frac{1}{p_u(1)}\left(\frac{dp_u(1)}{du}\right)^2 + \frac{1}{p_u(-1)}\left(\frac{dp_u(-1)}{du}\right)^2$$

$$= \frac{\cos(u)^2}{2}\left(\frac{1}{1+\sin(u)} + \frac{1}{1-\sin(u)}\right) = 1$$

Risk of maximum likelihood estimator $\hat{u}_n := \arcsin(n_+ - n_-)$

$$\mathbb{E}[(u-\hat{u})^2] \approx \frac{1}{n} I(u)^{-1} = \frac{1}{n}$$

Example: estimation of a pure spin state

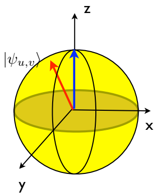


Problem

Estimate rotation paramaters (u, v) in 2D model

$$|\psi_{u,v}\rangle := \exp\left(\frac{iu\sigma_x - v\sigma_y}{2}\right)|\uparrow\rangle$$

Example: estimation of a pure spin state



Problem

Estimate rotation paramaters (u, v) in 2D model

$$|\psi_{u,v}\rangle := \exp\left(\frac{iu\sigma_x - v\sigma_y}{2}\right)|\uparrow\rangle$$

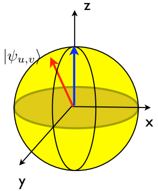
Two step adaptive procedure

1. Measure $\tilde{n} \ll n$ systems and obtain a preliminary estimator (e.g.$|\uparrow\rangle$)
2. Measure the orthogonal directions σ_x and σ_y, on $(n - \tilde{n})/2$ systems

Total Fisher information matrix at $(u, v) = (0, 0)$

$$I_n((0,0)) := \frac{n}{2}(I^{(x)} + I^{(y)}) = \frac{n}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \frac{n}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{n}{2}\mathbf{1}$$

Example: estimation of a pure spin state



Problem

Estimate rotation paramaters (u, v) in 2D model

$$|\psi_{u,v}\rangle := \exp\left(\frac{iu\sigma_x - v\sigma_y}{2}\right)|\uparrow\rangle$$

Two step adaptive procedure

1. Measure $\tilde{n} \ll n$ systems and obtain a preliminary estimator (e.g.$|\uparrow\rangle$)
2. Measure the orthogonal directions σ_x and σ_y, on $(n - \tilde{n})/2$ systems

Total Fisher information matrix at $(u, v) = (0, 0)$

$$I_n((0,0)) := \frac{n}{2}(I^{(x)} + I^{(y)}) = \frac{n}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \frac{n}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{n}{2}\mathbf{1}$$

Risk of estimator $(\hat{u}_n, \hat{v}_n) = \left((n_+^{(y)} - n_+^{(y)}), (n_+^{(x)} - n_-^{(x)})\right)$

$$\mathbb{E}[\|\vec{r}_{u,v} - \vec{r}_{\hat{u}_n,\hat{v}_n}\|^2] = \mathbb{E}[(u - \hat{u}_n)^2 + (v - \hat{v}_n)^2] \approx \mathrm{Tr}(I_n(0,0)^{-1}) = \frac{4}{n}$$

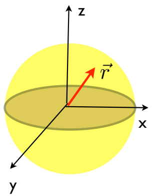
Can we extract more statistical information with other measurements?

Example: estimation of a mixed spin state

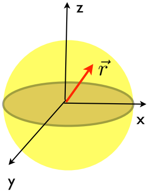
Problem

Estimate $\vec{r}$, given n quantum spins prepared in state

$$\rho_{\vec{r}} := \frac{1}{2}\begin{pmatrix} 1 + r_z & r_x - i r_y \\ r_x + i r_y & 1 - r_z \end{pmatrix} = \frac{1}{2}\left(\mathbf{1} + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z\right)$$



Example: estimation of a mixed spin state



Problem

Estimate $\vec{r}$, given n quantum spins prepared in state

$$\rho_{\vec{r}} := \frac{1}{2}\begin{pmatrix} 1 + r_z & r_x - i r_y \\ r_x + i r_y & 1 - r_z \end{pmatrix} = \frac{1}{2}\left(\mathbf{1} + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z\right)$$

Two step adaptive procedure

1. Measure $\tilde{n} \ll n$ systems and obtain a preliminary estimator, e.g.

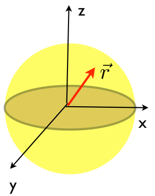
$$\rho_0 = \rho_{\vec{r}_0} := \frac{1 + r_0}{2}|\uparrow\rangle\langle\uparrow| + \frac{1 - r_0}{2}|\downarrow\rangle\langle\downarrow|$$

2. Estimate (r_x, r_y, r_z) by measuring $(\sigma_x, \sigma_y, \sigma_z)$ separately on $\left(\frac{\lambda(n-\tilde{n})}{2}, \frac{\lambda(n-\tilde{n})}{2}, (1-\lambda)(n-\tilde{n})\right)$ systems

Example: estimation of a mixed spin state

Problem

Estimate $\vec{r}$, given n quantum spins prepared in state



$$\rho_{\vec{r}} := \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix} = \frac{1}{2}\left(\mathbf{1} + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z\right)$$

Two step adaptive procedure

1. Measure $\tilde{n} \ll n$ systems and obtain a preliminary estimator, e.g.

$$\rho_0 = \rho_{\vec{r}_0} := \frac{1+r_0}{2}|\uparrow\rangle\langle\uparrow| + \frac{1-r_0}{2}|\downarrow\rangle\langle\downarrow|$$

2. Estimate (r_x, r_y, r_z) by measuring $(\sigma_x, \sigma_y, \sigma_z)$ separately on $\left(\frac{\lambda(n-\tilde{n})}{2}, \frac{\lambda(n-\tilde{n})}{2}, (1-\lambda)(n-\tilde{n})\right)$ systems

Risk for the optimal choice of λ

$$\mathbb{E}[\|\vec{r} - \hat{\vec{r}}_n\|^2] \approx \mathrm{Tr}(I_n(\vec{r}_0)^{-1}) = \frac{(2+\sqrt{1-r^2})^2}{n} < \frac{3(3-r^2)}{n}$$

Can we extract more statistical information with other measurements?

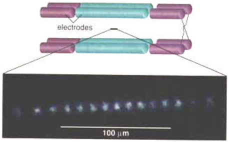
Outlook

- For a given repeated measurement we can use classical asymptotic theory to compute asymptotic rates of convergence and error bars
- Adaptive (separate) measurements perform better than fixed design ones
- Joint measurements perform better than separate ones for multi-dimensional models with mixed states (see second lecture)

3. The 8 (14) ions experiment

- Measurement and statistical model
- Fisher information and Cramér-Rao bound
- Asymptotic error for pure states
- The Bayesian information criterion (BIC)
- Behaviour of the Cramér-Rao bound with the rank
- Estimation error in models with higher rank

Quantum tomography for trapped ions



[Häffner *et al*, Nature 2005]

Goal: prepare a W (entangled) state of several (4 to 8) ions

Validation: statistical 'reconstruction' of the quantum state $\rho \in M(\mathbb{C}^{2^k})$

- $4^8 - 1 = 65\,535$ parameters to estimate (8 ions)
- $3^8 \times 100 = 656\,100$ repeated measurements
- 10 hours measurement time
- weeks of computer time ('maximum likelihood')
- fidelity between estimator and target state between 0.85 and 0.72

Measurement procedure and statistical model

All measurements are performed on independent identically prepared states $\rho \in M(\mathbb{C}^{2^k})$

1. For each ion choose a spin direction to measure $\sigma_d \in \{\sigma_x, \sigma_y, \sigma_z\}$
2. measure each qubit and obtain outcome $\mathbf{s} := (s_1, \dots, s_k) \in \{1, -1\}^k$

$$\mathbb{P}_\rho(\mathbf{s}|\mathbf{d}) = \mathbb{P}_\rho(s_1, \dots, s_k | \sigma_{d_1}, \dots, \sigma_{d_k}) = \left\langle e^{s_1}_{d_1} \otimes \cdots \otimes e^{s_k}_{d_k} \middle| \rho \middle| e^{s_1}_{d_1} \otimes \cdots \otimes e^{s_k}_{d_k} \right\rangle$$

3. Repeat 100 times and collect counts of outcomes $\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}$

$$\mathbb{P}_\rho(\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}) = \frac{100!}{\prod_s N_{\mathbf{s},\mathbf{d}}!} \prod_{\mathbf{s}} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d})^{N_{\mathbf{s},\mathbf{d}}}$$

4. Repeat over all 3^k choices of measurement set-ups

Total set of $3^k \times 2^k \gg 4^k$ projections is highly overcomplete in $M(\mathbb{C}^{2^k})$!

Measurement data

- 3^k columns of length 2^k
- one column for each measurement setting
- each column contains the counts of the 2^k possible outcomes totalling 100
- frequencies of outcomes are bad estimates of probabilities, but overall info is high

1	2	11	11	11	21	5	16	21	19	11	16	2	26	15	5
2	19	10	6	15	4	22	10	3	12	8	16	18	5	14	16
3	30	12	15	9	10	18	14	3	6	11	4	4	2	1	5
4	0	4	15	10	17	2	4	14	13	0	4	8	5	1	3
5	21	13	12	7	6	5	14	12	8	12	7	19	3	8	3
6	1	12	14	0	1	1	0	6	6	12	8	2	6	2	7
7	1	2	0	19	7	12	14	6	7	14	7	9	23	15	34
8	0	1	1	0	4	8	0	6	6	0	7	12	4	15	5
9	21	17	8	10	7	7	14	9	8	15	6	9	6	3	0
10	2	16	15	0	12	9	0	3	4	1	7	3	0	4	6
11	0	0	1	17	9	2	14	12	7	0	1	0	5	5	2
12	1	1	1	0	2	8	0	4	3	0	1	0	0	3	1
13	1	0	1	1	0	0	0	0	0	14	9	7	6	2	4
14	0	1	0	0	0	1	0	0	1	1	5	6	0	2	2
15	1	0	0	1	0	0	0	0	0	1	2	0	9	6	3
16	0	0	0	0	0	0	0	1	0	0	0	1	0	4	4

[Data set 4 ions (from H. Häffner)]

Questions

- Why did it work ? Would it work for a very mixed state as well?
- What is the structure of the data? Are we in an asymptotic regime ?
- Are there other less expensive estimation methods ?

Outline

- Measurement and statistical model
- Fisher information and Cramér-Rao bound
- Asymptotic error for pure states
- The Bayesian information criterion (BIC)
- Behaviour of the Cramér-Rao bound with the rank
- Estimation error in models with higher rank

Asymptotics in estimation of biased coin

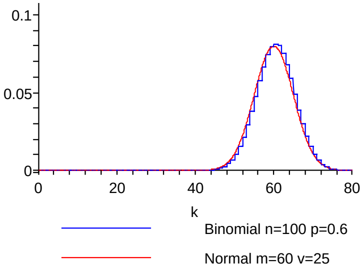
- $X_1, \ldots X_n$ i.i.d. with $\mathbb{P}[X_i = 1] = \theta$ and $\mathbb{P}[X_i = 0] = 1 - \theta$
- Estimator $\hat{\theta}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$

Asymptotics in estimation of biased coin

- $X_1, \dots X_n$ i.i.d. with $\mathbb{P}[X_i = 1] = \theta$ and $\mathbb{P}[X_i = 0] = 1 - \theta$
- Estimator $\hat{\theta}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$
- Central Limit Theorem $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1 - \theta))$

Asymptotics in estimation of biased coin

- $X_1, \ldots X_n$ i.i.d. with $\mathbb{P}[X_i = 1] = \theta$ and $\mathbb{P}[X_i = 0] = 1 - \theta$
- Estimator $\hat{\theta}_n := \frac{1}{n}\sum_{i=1}^n X_i$
- Central Limit Theorem $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1-\theta))$



Fisher information, Cramér-Rao bound and asymptotic normality

Let $X_1, \ldots, X_n$ be i.i.d. data with probability distribution $\mathbb{P}_\theta$ and $\theta \in \mathbb{R}^p$

- Cramér-Rao bound
  for every estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ which is unbiased, i.e. $\mathbb{E}(\hat{\theta}_n) = \theta$

$$\mathbb{E}_\theta \left[(\hat{\theta}_n - \theta)^T (\hat{\theta}_n - \theta) \right] \geq (nI_\theta)^{-1}$$

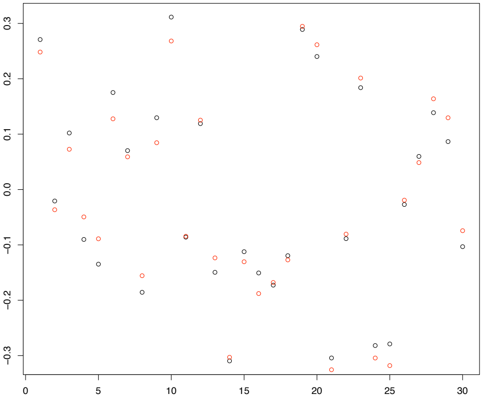
  where I^θ is the Fisher information matrix

$$I(\theta)_{i,j} := \int \frac{\partial p_\theta(x)}{\partial \theta_i} \frac{\partial p_\theta(x)}{\partial \theta_i} p_\theta(x) dx$$

- "good" estimators (e.g. max.lik. under certain conditions) are asymptotically normal

$$\sqrt{n}(\hat{\theta}_n - \theta) \approx N(0, I^{-1}(\theta))$$

Maximum likelihood estimation for pure states

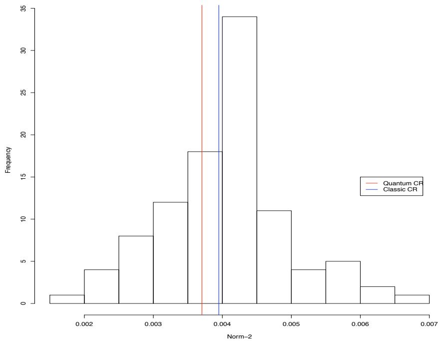


Parameters of true (black) versus estimated (red) pure state of 4 ions

Maximum likelihood estimator $\hat{\rho}_{\rm ml} = \hat{\rho}_{\rm ml}(\{N_{\mathbf{s},\mathbf{d}}\})$

$$\hat{\rho}_{\rm ml} := \arg\max_{\tau} \prod_{\mathbf{d}} \mathbb{P}_{\tau}(\{N_{\mathbf{s},\mathbf{d}}\})$$

Maximum likelihood estimation for pure states



Histogram of $\|\hat{\rho}_{ml} - \rho\|^2$ for a pure state $\rho \in M(\mathbb{C}^{2^4})$ (100 repetitions)

- Very good agreement with asymptotic theory
  - median very close to the Cramer-Rao bound (blue line)
- Measurement is very close to optimal
  - Quantum CR bound $30/(100 * 3^4) = 0.0037$ (red line) slightly smaller than classical CR

Outline

- Measurement and statistical model
- Fisher information and Cramér-Rao bound
- Asymptotic error for pure states
- The Bayesian information criterion (BIC)
- Behaviour of the Cramér-Rao bound with the rank
- Estimation error in models with higher rank

Choosing the rank of the state by BIC

If state is **not** known to be pure, can we estimate it without doing ML over **all** states?

Choosing the rank of the state by BIC

If state is **not** known to be pure, can we estimate it without doing ML over **all** states?

1. Perform separate ML over states of rank $r = 1, 2, 3, \ldots$ to obtain $\hat{\rho}^{(1)}_{\mathrm{ml}}, \hat{\rho}^{(2)}_{\mathrm{ml}}, \hat{\rho}^{(3)}_{\mathrm{ml}} \ldots$

Choosing the rank of the state by BIC

If state is **not** known to be pure, can we estimate it without doing ML over **all** states?

1. Perform separate ML over states of rank $r = 1, 2, 3, ...$ to obtain $\hat{\rho}^{(1)}_{\mathrm{ml}}, \hat{\rho}^{(2)}_{\mathrm{ml}}, \hat{\rho}^{(3)}_{\mathrm{ml}} \ldots$

2. Choose the rank r which minimises the Bayesian information criterion (BIC):

$$
\begin{aligned}
\mathrm{BIC}(r) &= -2 \log \mathbb{P}_{\hat{\rho}^{(r)}_{\mathrm{ml}}}(\mathrm{DATA}) + \sharp\mathrm{parameters}(\hat{\rho}^{(r)}_{\mathrm{ml}}) * \log n \\
&= -2 \log \mathbb{P}_{\hat{\rho}^{(r)}_{\mathrm{ml}}}(\{N_{\mathbf{s},\mathbf{d}}\}) + (2r * 2^k - r^2 - 1) * \log n
\end{aligned}
$$

Theoretical motivation:

In a Bayesian set-up the states are drawn by first choosing the rank according to a prior $\{\pi(r)\}$, followed by choosing a state of rank r from some distribution.

Then the rank with the highest posterior probability is selected

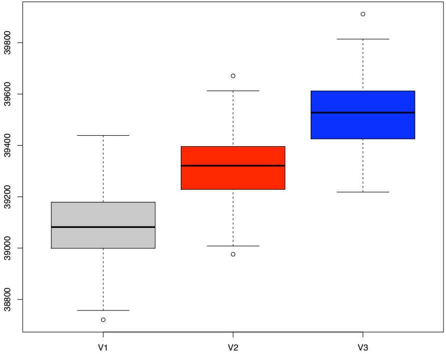
The BIC is an asymptotic approximation to the log of the posterior likelihood.

BIC performance

BIC chosen rank

true rank	1	2	3
1	99	0	1
2	0	90	10
3	0	6	94

BIC performance in 100 repetitions from states of rank 1,2,3



Boxplot of BIC values for a pure state

Outline

- Measurement and statistical model
- Fisher information and Cramér-Rao bound
- Asymptotic error for pure states
- The Bayesian information criterion (BIC)
- Behaviour of the Cramér-Rao bound with the rank
- Estimation error in models with higher rank

Classical Cramér-Rao bound vs Holevo bound

Quantum statistical model of rank r states $\mathcal{Q}_r := \{\rho_\theta : \theta \in \Theta(r) \subset \mathbb{R}^{\dim(r)}\}$

Classical Cramér-Rao bound

$$n\mathbb{E}\|\rho - \hat{\rho}_n\|_2^2 = n\mathbb{E}\|\rho_\theta - \rho_{\hat{\theta}_n}\|_2^2 \geq \mathrm{Tr}(G(\theta)I(\theta)^{-1})$$

- $I(\theta)$ is the $\dim(r) \times \dim(r)$ (measurement dependent) Fisher info. matrix
- $G(\theta)$ is the $\dim(r) \times \dim(r)$ matrix of the quadratic approximation of loss

$$\|\rho - \hat{\rho}\|_2^2 = \|\rho_\theta - \rho_{\hat{\theta}}\|_2^2 \approx (\theta - \hat{\theta})G(\theta)(\theta - \hat{\theta})^T$$

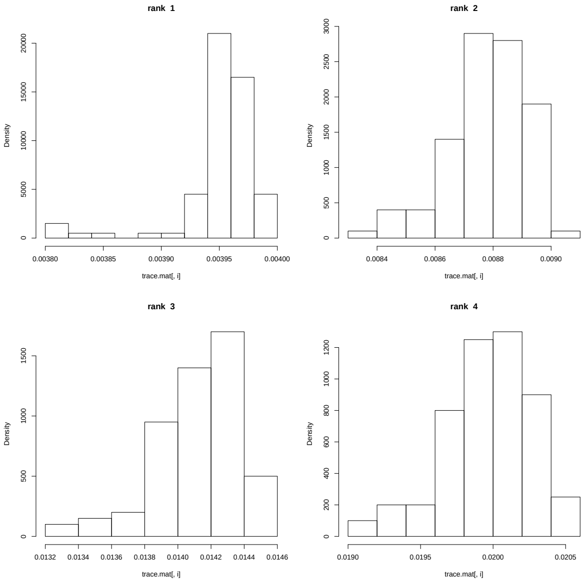
Quantum Holevo bound:

BEST measurement & estimator for a state ρ with eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_d$

$$n\mathbb{E}\|\rho - \hat{\rho}_n\|_2^2 \to \sum_{i=1}^{d} \mu_i(1-\mu_i) + 2\sum_{j<k} \mu_j \leq 2d + 1$$

does not increase with the rank!

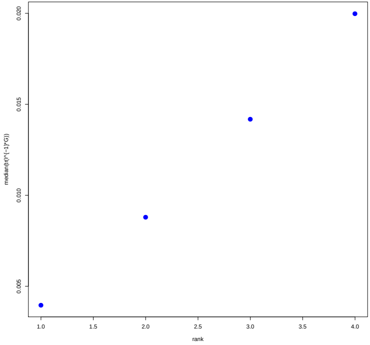
Error rates as function of the rank of true state



Histogram of CR bound for random states of different ranks

Error rates as function of the rank of true state

Mean square error for ions measurement appears to increase linearly with rank

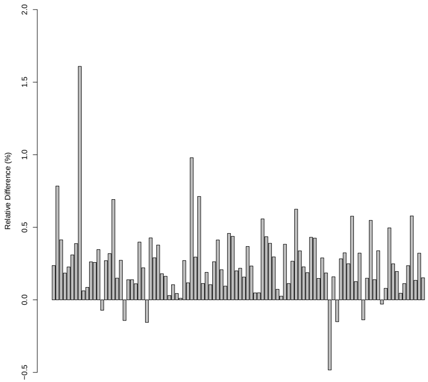


Median of of CR bound as function of rank

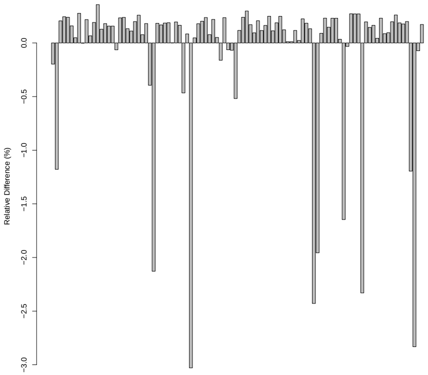
Outline

- Measurement and statistical model
- Fisher information and Cramér-Rao bound
- Asymptotic error for pure states
- The Bayesian information criterion (BIC)
- Behaviour of the Cramér-Rao bound with the rank
- Estimation error in models with higher rank

Relative difference in error for different ranks



Histogram of $1 - \frac{\|\rho - \hat{\rho}^{(1)}_{\mathrm{ml}}\|_2^2}{\|\rho - \hat{\rho}^{(2)}_{\mathrm{ml}}\|_2^2}$ for state of rank 1



Histogram of $1 - \frac{\|\rho - \hat{\rho}^{(3)}_{\mathrm{ml}}\|_2}{\|\rho - \hat{\rho}^{(2)}_{\mathrm{ml}}\|_2}$ for state of rank 2

Behaviour of the Cramér-Rao bound for nested models

- Successive inclusions of subspaces of states of given rank

$$\mathcal{S}(1) \subset \mathcal{S}(2) \subset ..$$

- True state $\rho = \rho_{\theta(r)} \in \mathcal{S}(r)$

- Estimate as a state of rank $r' \geq r$ i.e. $\rho = \rho_{\theta(r')}$

Claim:

$$\mathrm{Tr}\left(I(\theta^{(r)})^{-1}G(\theta^{(r)})\right) = \mathrm{Tr}\left(I(\theta^{(r')})^{-1}G(\theta^{(r')})\right)$$

Explanation:
Additional parameters represent eigenvectors for very small eigenvalues
Model can be seen as mixture of pure state models, with some very small probabilities

$$p(\theta_1, \theta_2) = \mu_1 p(\theta_1) + \mu_2 p(\theta_2), \qquad \mu_2 \ll 1$$

loss function (figure of merit) is insensitive to errors in parameters for small weight

$$D\left[(\theta_1, \theta_2), (\hat{\theta}_1, \hat{\theta}_2)\right] \approx D(\theta_1, \hat{\theta}_1)$$

Outlook

- Measurement was well within the asymptotic set-up
- For pure states the “8 ions measurement” peforms very close to optimal measurement
- For mixed states the MSE increases linearly with the rank
- BIC performs well in selecting the rank of true state;
- BIC provides faster estimation method than “global” ML
- MSE is not very sensitive to overestimating the rank of the state

Outline of the mini-course

1. Notions of statistical inference
2. Basics of quantum state estimation
3. The 8 (14) ions experiment
4. The quantum Cramér-Rao theory
5. Local asymptotic normality for i.i.d. quantum states
6. Local asymptotic normality for quantum Markov chains

4. The quantum Cramér-Rao theory

- The $L^2(\rho)$ Hilbert space
- The quantum Fisher-Helstrom information matrix
- Quantum Cramér-Rao bound
- The quantum Cramér-Rao bound is achievable for $\Theta \subset \mathbb{R}$
- (Non)-achievability of the quantum C.-R. bound for $\Theta \subset \mathbb{R}^k$ with $k > 1$
- The Holevo bound

Quantum Statistics pioneers

- Helstrom, Holevo, Belavkin, Yuen, Kennedy...
- Formulated and solved first quantum statistical decision problems
  - quantum statistical model $\mathcal{Q} = \{\rho_\theta \ : \ \theta \in \Theta\}$
  - decision problem (estimation, testing)
  - find optimal measurement (and estimator)
- Quantum Gaussian states, covariant families, state discrimination...
- Elements of a (purely) quantum statistical theory
  - Quantum Fisher Information
  - Quantum Cramér-Rao bound(s)
  - Holevo bound for quadratic risk
  - ...

Motivating questions

- How much statistical information can be extracted from a quantum model ?
- Is there a quantum analogue of asymptotic normality ?
- Is there a quantum analogue of likelihood ratio, sufficiency, ...

The $L^2(\rho)$ Hilbert space

Definition

Let ρ be a state on $\mathbb{C}^d$. We denote by $L^2_{\mathbb{R}}(\rho)$ the Hilbert space $(M(\mathbb{C}^d)_{sa}, \langle\cdot,\cdot\rangle_\rho)$ with inner product

$$\langle A, B\rangle_\rho := \mathrm{Tr}\,(\rho\, A \circ B)\,, \qquad A \circ B := \frac{1}{2}(AB + BA)$$

The $L^2(\rho)$ Hilbert space

Definition

Let ρ be a state on $\mathbb{C}^d$. We denote by $L^2_{\mathbb{R}}(\rho)$ the Hilbert space $(M(\mathbb{C}^d)_{sa}, \langle\cdot,\cdot\rangle_\rho)$ with inner product

$$\langle A, B\rangle_\rho := \mathrm{Tr}\left(\rho\, A \circ B\right), \qquad A \circ B := \frac{1}{2}(AB + BA)$$

Remark

- If $\mathrm{Tr}(\rho(A - B)^2) = 0$ then A and B correspond to the same vector in $L^2_{\mathbb{R}}(\rho)$. This identification is relevant when ρ is not full rank.
- For infinite dimensional spaces $L^2_{\mathbb{R}}(\rho)$ is defined as the completion of $\mathcal{B}(\mathcal{H})_{sa}$ with respect to $\langle\cdot,\cdot\rangle_\rho$. Each vector in $L^2_{\mathbb{R}}(\rho)$ can be identified with (the equivalence class of) a square summable operator w.r.t. ρ, i.e. unbounded symmetric linear operators satisfying

$$\sum \lambda_i \|Xe_i\|^2 < \infty$$

where $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$ is the spectral decomposition of ρ.

The quantum Fisher-Helstrom information matrix

Let $\{\rho_\theta : \theta \in \Theta\}$ be a parametric statistical model with $\rho_\theta \in M(\mathbb{C}^d)$ and $\Theta \subset \mathbb{R}^k$ open, and assume that $\theta \mapsto \rho_\theta$ is a differentiable function.

The quantum Fisher-Helstrom information matrix

Let $\{\rho_\theta : \theta \in \Theta\}$ be a parametric statistical model with $\rho_\theta \in M(\mathbb{C}^d)$ and $\Theta \subset \mathbb{R}^k$ open, and assume that $\theta \mapsto \rho_\theta$ is a differentiable function.

Definition

- The symmetric logarithmic derivative (s.l.d.) for the coordinate θ_i is the unique vector $\mathcal{L}_{\theta,i} \in L^2_{\mathbb{R}}(\rho_\theta)$ such that

$$\frac{\partial \rho_\theta}{\partial \theta_i} = \mathcal{L}_{\theta,i} \circ \rho_\theta$$

- The quantum Fisher-Helstrom information matrix is defined as

$$H(\theta)_{i,j} = \langle \mathcal{L}_{\theta,i}, \mathcal{L}_{\theta,j} \rangle_\theta = \mathrm{Tr}(\rho_\theta \mathcal{L}_{\theta,i} \circ \mathcal{L}_{\theta,j})$$

Let $\{\rho_\theta : \theta \in \Theta\}$ be a parametric statistical model with $\rho_\theta \in M(\mathbb{C}^d)$ and $\Theta \subset \mathbb{R}^k$ open, and assume that $\theta \mapsto \rho_\theta$ is a differentiable function.

Definition

- The symmetric logarithmic derivative (s.l.d.) for the coordinate θ_i is the unique vector $\mathcal{L}_{\theta,i} \in L^2_{\mathbb{R}}(\rho_\theta)$ such that

$$\frac{\partial \rho_\theta}{\partial \theta_i} = \mathcal{L}_{\theta,i} \circ \rho_\theta$$

- The quantum Fisher-Helstrom information matrix is defined as

$$H(\theta)_{i,j} = \langle \mathcal{L}_{\theta,i}, \mathcal{L}_{\theta,j} \rangle_\theta = \mathrm{Tr}(\rho_\theta \mathcal{L}_{\theta,i} \circ \mathcal{L}_{\theta,j})$$

Remark

- For infinite dimensional spaces we need to assume that the linear map $\varphi_i : A \mapsto \mathrm{Tr}(\partial\rho_\theta/\partial\theta_i A)$ can be extended to a continuous funct. on $L^2_{\mathbb{R}}(\rho_\theta)$. The s.l.d. is then defined by $\varphi_i(A) = \langle A, \mathcal{L}_{\theta,i} \rangle_\theta$ (cf. Riesz Theorem)

- When $\{\rho_\theta : \theta \in \Theta\}$ form a commuting family, the s.l.d. $\mathcal{L}_{\theta,i}$ can be identified with the classical score function $\dot{\ell}_{\theta,i} = \partial \log p_\theta / \partial \theta_i$

Properties of the quantum Fisher-Helstrom information matrix

- $H(\theta)$ is a real positive definite matrix

Properties of the quantum Fisher-Helstrom information matrix

- $H(\theta)$ is a real positive definite matrix
- additivity: if $\rho_\theta = \rho_\theta^{(1)} \otimes \rho_\theta^{(2)}$ then $H(\theta) = H(\theta)^{(1)} + H(\theta)^{(2)}$ (exercise)

Properties of the quantum Fisher-Helstrom information matrix

- $H(\theta)$ is a real positive definite matrix
- additivity: if $\rho_\theta = \rho_\theta^{(1)} \otimes \rho_\theta^{(2)}$ then $H(\theta) = H(\theta)^{(1)} + H(\theta)^{(2)}$ (exercise)
- The Bures (fidelity) distance between two states is defined by

$$b(\rho_1, \rho_2)^2 := 2(1 - \mathrm{Tr}(\sqrt{\rho_1^{1/2} \rho_2 \rho_1^{1/2}}))$$

  Infinitesimally, the Bures distance is given by

$$b(\rho_\theta, \rho_{\theta+d\theta})^2 = \frac{1}{4} H(\theta)(d\theta)^2 + o((d\theta)^2)$$

Properties of the quantum Fisher-Helstrom information matrix

- $H(\theta)$ is a real positive definite matrix
- additivity: if $\rho_\theta = \rho_\theta^{(1)} \otimes \rho_\theta^{(2)}$ then $H(\theta) = H(\theta)^{(1)} + H(\theta)^{(2)}$ (exercise)
- The Bures (fidelity) distance between two states is defined by

$$b(\rho_1, \rho_2)^2 := 2(1 - \mathrm{Tr}(\sqrt{\rho_1^{1/2}\rho_2\rho_1^{1/2}}))$$

  Infinitesimally, the Bures distance is given by

$$b(\rho_\theta, \rho_{\theta+d\theta})^2 = \frac{1}{4}H(\theta)(d\theta)^2 + o((d\theta)^2)$$

- contractivity: let $C : \mathcal{T}_1(\mathcal{H}) \to \mathcal{T}_1(\mathcal{K})$ be a quantum channel (completely positive, trace preserving linear map). Let $\tau_\theta := C(\rho_\theta)$ be the quantum model obtained by applying the 'quantum randomisation' C to ρ_θ. Then

$$b(\rho_{\theta_1}, \rho_{\theta_2}) \geq b(\tau_{\theta_1}, \tau_{\theta_2}), \qquad \text{and} \qquad H(\rho_\theta) \geq H(\tau_\theta)$$

Properties of the quantum Fisher-Helstrom information matrix

- $H(\theta)$ is a real positive definite matrix
- additivity: if $\rho_\theta = \rho_\theta^{(1)} \otimes \rho_\theta^{(2)}$ then $H(\theta) = H(\theta)^{(1)} + H(\theta)^{(2)}$ (exercise)
- The Bures (fidelity) distance between two states is defined by

  $$b(\rho_1, \rho_2)^2 := 2(1 - \mathrm{Tr}(\sqrt{\rho_1^{1/2} \rho_2 \rho_1^{1/2}}))$$

  Infinitesimally, the Bures distance is given by

  $$b(\rho_\theta, \rho_{\theta+d\theta})^2 = \frac{1}{4} H(\theta)(d\theta)^2 + o((d\theta)^2)$$

- contractivity: let $C : \mathcal{T}_1(\mathcal{H}) \to \mathcal{T}_1(\mathcal{K})$ be a quantum channel (completely positive, trace preserving linear map). Let $\tau_\theta := C(\rho_\theta)$ be the quantum model obtained by applying the 'quantum randomisation' C to ρ_θ. Then

  $$b(\rho_{\theta_1}, \rho_{\theta_2}) \geq b(\tau_{\theta_1}, \tau_{\theta_2}), \qquad \text{and} \qquad H(\rho_\theta) \geq H(\tau_\theta)$$

- unlike the classical case, H is not the unique contractive metric. Such metrics are in one-to-one correspondence with operator monotone functions $f : \mathbb{R}_+ \to \mathbb{R}$ (i.e. $f(A) \geq f(B)$ for all $A \geq B \geq 0$ in $\mathcal{B}(\mathcal{H})$) satisfying $f(t) = tf(t^{-1})$ and $f(1) = 1$

Reference: D. Petz, Linear Algebra Appl. **244** 81-96 (1996)

Quantum Cramér-Rao bound (I)

Theorem

Let $\mathcal{Q} := \{\rho_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ *be a quantum statistical model with quantum Fisher-Helstrom, information matrix* $H(\theta)$*.*

Quantum Cramér-Rao bound (I)

Theorem

Let $\mathcal{Q} := \{\rho_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a quantum statistical model with quantum Fisher-Helstrom, information matrix $H(\theta)$.

Let M be a measurement with outcomes in $\{1, \ldots, k\}$ and let $I_M(\theta)$ be the Fisher information matrix of the classical statistical model $\mathcal{P}_M := \{\mathbb{P}_\theta^{(M)} : \theta \in \Theta\}$.

Quantum Cramér-Rao bound (I)

Theorem

Let $\mathcal{Q} := \{\rho_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ *be a quantum statistical model with quantum Fisher-Helstrom, information matrix* $H(\theta)$.

Let M *be a measurement with outcomes in* $\{1, \ldots, k\}$ *and let* $I_M(\theta)$ *be the Fisher information matrix of the classical statistical model* $\mathcal{P}_M := \{\mathbb{P}_\theta^{(M)} : \theta \in \Theta\}$.

Then the matrix inequality holds

$$I_M(\theta) \leq H(\theta)$$

and in particular, for any unbiased estimator $\hat{\theta}$ *of* θ *we have*

$$Var(\hat{\theta}) \geq I_M(\theta)^{-1} \geq H(\theta)^{-1}$$

Quantum Cramér-Rao bound (I)

Theorem

Let $\mathcal{Q} := \{\rho_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ *be a quantum statistical model with quantum Fisher-Helstrom, information matrix* $H(\theta)$*.*

Let M *be a measurement with outcomes in* $\{1, \ldots, k\}$ *and let* $I_M(\theta)$ *be the Fisher information matrix of the classical statistical model* $\mathcal{P}_M := \{\mathbb{P}_\theta^{(M)} : \theta \in \Theta\}$ *.*

Then the matrix inequality holds

$$I_M(\theta) \leq H(\theta)$$

and in particular, for any unbiased estimator $\hat{\theta}$ *of* θ *we have*

$$Var(\hat{\theta}) \geq I_M(\theta)^{-1} \geq H(\theta)^{-1}$$

Remark

- In the last display, the left inequality is the 'classical' Cramér-Rao.
- the right inequality follows from applying the operator monotone function $f(x) = x^{-1}$ to the previous inequality $I_M(\theta) \leq H(\theta)$.
- A function is called operator monotone if $f(A) \leq f(B)$ for all bounded operators satisfying $0 \leq A \leq B$. Not all monotone functions are operator

Example: unitary family

Let $\rho_\theta := \exp(-i\theta K)\rho\exp(i\theta K)$ with $\rho = \sum_i \lambda_i |i\rangle\langle i|$ and $\theta \in \mathbb{R}$.

- Symmetric logarithmic derivative

$$\left.\frac{d\rho_\theta}{d\theta}\right|_{\theta=0} = -i[K,\rho] = \rho \circ \mathcal{L}$$

- Solution

$$\langle i|\mathcal{L}|j\rangle = \frac{2i(\lambda_i - \lambda_j)}{\lambda_i + \lambda_j}\langle i|K|j\rangle$$

- Quantum Fisher information $H(\theta) = H$

$$H = \mathrm{Tr}(\rho\mathcal{L}^2) = 4\sum_{ij}\lambda_i\left(\frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j}\right)^2 |\langle i|K|j\rangle|^2 \xrightarrow{\text{pure states}} H = 4\langle\psi|K^2|\psi\rangle$$

Example (exercise)

Let $\rho_{\mathbf{r}}$ be the qubit state with Bloch vector $\mathbf{r}$ represented in polar coordinates $\mathbf{r} \leftrightarrow (r, \theta, \phi)$

$$\rho_{\mathbf{r}} = \frac{1}{2} \begin{pmatrix} 1 + r\cos\theta & r\sin\theta e^{-i\phi} \\ r\sin\theta e^{-i\phi} & 1 - r\cos\theta \end{pmatrix} = \frac{1}{2}(\mathbf{1} + \mathbf{r}\sigma)$$

Symmetric logarithmic derivatives

$$\frac{\partial \rho_{\mathbf{r}}}{\partial r} = \mathcal{L}_{\mathbf{r},r} \circ \rho_{\mathbf{r}}, \qquad \frac{\partial \rho_{\theta}}{\partial \theta} = \mathcal{L}_{\mathbf{r},\theta} \circ \rho_{\mathbf{r}}, \qquad \frac{\partial \rho_{\mathbf{r}}}{\partial \phi} = \mathcal{L}_{\mathbf{r},\phi} \circ \rho_{\phi}$$

with solutions

$$\mathcal{L}_r = \frac{1}{1+r}(\mathbf{1} + \mathbf{r}\sigma/r), \qquad \mathcal{L}_\theta = \frac{\partial \mathbf{r}}{\partial \theta}\sigma, \qquad \mathcal{L}_\phi = \frac{\partial \mathbf{r}}{\partial \phi}\sigma.$$

Quantum Fisher-Helstrom information matrix

$$H(\mathbf{r}) = \begin{pmatrix} \frac{1}{1-r^2} & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin\theta^2 \end{pmatrix}$$

The proof [Braunstein and Caves (1994)]

Consider θ one dimensional. General case is left as an exercise
Let $M = (M_1, \ldots, M_k)$ be the POVM.

Differentiating $p_\theta(i) = \mathrm{Tr}(\rho_\theta M_i)$ and using $d\rho_\theta / d\theta = \mathcal{L}_\theta \circ \rho_\theta$ we get

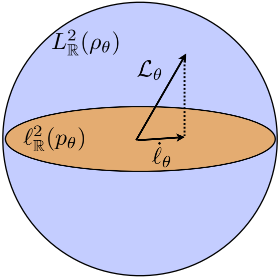
$$\frac{dp_\theta(i)}{d\theta} = \mathrm{Tr}(\mathcal{L}_\theta \circ \rho_\theta M_i) = \mathbf{Re}\mathrm{Tr}(\rho_\theta \mathcal{L}_\theta M_i)$$

Then with $\mathcal{I} := \{i : p_\theta(i) \neq 0\}$ we have the inequalites

$$\begin{aligned}
I_M(\theta) &= \sum_{i\in\mathcal{I}} p_\theta(i)^{-1} \left(\mathbf{Re}\mathrm{Tr}(\rho_\theta \mathcal{L}_\theta M_i)\right)^2 \\
&\leq \sum_{i\in\mathcal{I}} p_\theta(i)^{-1} \left|\mathrm{Tr}(\rho_\theta \mathcal{L}_\theta M_i)\right|^2 \\
&= \sum_{i\in\mathcal{I}} \mathrm{Tr}(\rho_\theta M_i)^{-1} \left|\mathrm{Tr}((M_i^{1/2}\rho_\theta^{1/2})^* M_i^{1/2} \mathcal{L}_\theta \rho_\theta^{1/2})\right|^2 \\
&\leq \sum_{i\in\mathcal{I}} \mathrm{Tr}(M_i \mathcal{L}_\theta \rho_\theta \mathcal{L}_\theta) \leq \sum_{i=1}^{k} \mathrm{Tr}(M_i \mathcal{L}_\theta \rho_\theta \mathcal{L}_\theta) \\
&= H(\theta)
\end{aligned}$$

where we used Cauchy-Schwarz in the second inequality

The algebraic proof: geometric idea



Both classical and quantum Fisher informations are equal to the square lengths of Hilbert space vectors

$I_M(\theta) = \|\dot{\ell}_\theta\|^2$ with $\dot{\ell}_\theta \in \ell^2(p_\theta)$
$H(\theta) = \|\mathcal{L}_\theta\|^2$ with $\mathcal{L}_\theta \in L^2(\rho_\theta)$

With the appropriate embedings, $\dot{\ell}_\theta$ is the projection of $\mathcal{L}_\theta$ onto $\ell^2(p_\theta)$, hence

$$I_M(\theta) \leq H(\theta)$$

See Quantum Statistics notes at http://maths.dept.shef.ac.uk/magic/course.php?id=181 for the full proof

The quantum Cramér-Rao bound is asymptotically achievable for $\Theta \subset \mathbb{R}$

- Let L be the result of measuring $\mathcal{L}_{\theta_0}$
- For $\theta = \theta_0$

$$\mathbb{E}_{\theta_0}(L) = \mathrm{Tr}(\rho_{\theta_0}\mathcal{L}_{\theta_0}) = 0, \qquad \mathrm{Var}_{\theta_0}(L) = \mathrm{Tr}(\rho_{\theta_0}\mathcal{L}_{\theta_0}^2) = H(\theta_0)$$

The quantum Cramér-Rao bound is asymptotically achievable for $\Theta \subset \mathbb{R}$

- Let L be the result of measuring $\mathcal{L}_{\theta_0}$
- For $\theta = \theta_0$

$$\mathbb{E}_{\theta_0}(L) = \mathrm{Tr}(\rho_{\theta_0}\mathcal{L}_{\theta_0}) = 0, \qquad \mathrm{Var}_{\theta_0}(L) = \mathrm{Tr}(\rho_{\theta_0}\mathcal{L}^2_{\theta_0}) = H(\theta_0)$$

- Estimator

$$\hat{\theta} := \theta_0 + \frac{L}{H(\theta_0)}$$

is locally unbiased around θ_0 since

$$\begin{aligned}
\mathbb{E}_\theta(\hat{\theta}) &= \theta_0 + \frac{\mathrm{Tr}(\rho_\theta\mathcal{L}_{\theta_0})}{H(\theta_0)} = \theta_0 + d\theta\frac{\mathrm{Tr}(\frac{d\rho_\theta}{d\theta}\mathcal{L}_{\theta_0})}{H(\theta_0)} + o(d\theta) \\
&= \theta_0 + d\theta\frac{\mathrm{Tr}(\rho_{\theta_0}\mathcal{L}^2_{\theta_0})}{H(\theta_0)} + o(d\theta) = \theta + o(d\theta)
\end{aligned}$$

The quantum Cramér-Rao bound is asymptotically achievable for $\Theta \subset \mathbb{R}$

- Let L be the result of measuring $\mathcal{L}_{\theta_0}$
- For $\theta = \theta_0$

$$\mathbb{E}_{\theta_0}(L) = \mathrm{Tr}(\rho_{\theta_0}\mathcal{L}_{\theta_0}) = 0, \qquad \mathrm{Var}_{\theta_0}(L) = \mathrm{Tr}(\rho_{\theta_0}\mathcal{L}_{\theta_0}^2) = H(\theta_0)$$

- Estimator

$$\hat{\theta} := \theta_0 + \frac{L}{H(\theta_0)}$$

is locally unbiased around θ_0 since

$$\begin{aligned}\mathbb{E}_\theta(\hat{\theta}) &= \theta_0 + \frac{\mathrm{Tr}(\rho_\theta \mathcal{L}_{\theta_0})}{H(\theta_0)} = \theta_0 + d\theta \frac{\mathrm{Tr}(\frac{d\rho_\theta}{d\theta}\mathcal{L}_{\theta_0})}{H(\theta_0)} + o(d\theta) \\ &= \theta_0 + d\theta \frac{\mathrm{Tr}(\rho_{\theta_0}\mathcal{L}_{\theta_0}^2)}{H(\theta_0)} + o(d\theta) = \theta + o(d\theta)\end{aligned}$$

and its variance is

$$\mathrm{Var}(\hat{\theta}) = \frac{\mathrm{Var}(L)}{H(\theta_0)^2} = H(\theta_0)^{-1}$$

The quantum Cramér-Rao bound is asymptotically achievable for $\Theta \subset \mathbb{R}$

Rigourous argument in the asymptotic framework using an adaptive procedure:

1. measure fraction $\tilde{n} \ll n$ of systems to obtain rough estimator θ_0
2. measure $\mathcal{L}^{(n)}_{\theta_0} := \mathcal{L}_{\theta_0} \otimes \mathbf{1} \otimes \cdots \otimes \mathbf{1} + \cdots + \mathbf{1} \otimes \mathbf{1} \otimes \cdots \otimes \mathcal{L}_{\theta_0}$
3. set $\hat{\theta}_n := \theta_0 + \mathbf{L}^{(n)}_{\theta_0} / H(\theta_0)$

The estimator is asymptotically efficient

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, H(\theta)^{-1})$$

Achievability of the quantum Cramér-Rao bound for $\Theta \subset \mathbb{R}^k$ with $k > 1$

- Bound is achievable iff

$$\mathrm{Tr}(\rho_\theta[\mathcal{L}_{\theta,j}, \mathcal{L}_{\theta,i}]) = 0, \qquad \forall 1 \leq i, j \leq k$$

- Bound is sharp

$$\mathrm{Cov}(\hat{\theta}) \geq K^{-1}(\theta), \quad \forall \text{ unbiased } M \quad \Longrightarrow \quad H(\theta)^{-1} \geq K^{-1}(\theta)$$

- Bound is not achievable in e.g., 2d-qubit rotation model, gaussian displacement

- What is a 'good estimator' in this case?
  - Trade-off between estimation of different coordinates
  - Optimal measurement depends on loss function

The Holevo bound for quadratic risk

Let $\mathcal{Q} = \{\rho_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a quantum statistical model on $\mathcal{H}$ and let $W(\hat{\theta}, \theta)$ be a quadratic loss function, i.e.

$$W(\hat{\theta}, \theta) = \sum_{i,j} (\hat{\theta}_i - \theta_i) G_{ij} (\hat{\theta}_j - \theta_j) = (\hat{\theta} - \theta) G (\hat{\theta} - \theta)^T$$

The Holevo bound for quadratic risk

Let $\mathcal{Q} = \{\rho_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ be a quantum statistical model on $\mathcal{H}$ and let $W(\hat{\theta}, \theta)$ be a quadratic loss function, i.e.

$$W(\hat{\theta}, \theta) = \sum_{i,j} (\hat{\theta}_i - \theta_i) G_{ij} (\hat{\theta}_j - \theta_j) = (\hat{\theta} - \theta) G (\hat{\theta} - \theta)^T$$

Theorem (Holevo bound)

For any measurement M with unbiased outcome $\hat{\theta}$ the following bound holds:

$$\mathrm{Tr}(G Var(\hat{\theta})) \geq \inf_{\mathbf{X}_\theta} \left\{ \mathrm{Tr}\left(\sqrt{G} \mathrm{Re}(Z(\mathbf{X}_\theta)) \sqrt{G} \right) + \mathrm{Tr}\left(\left| \sqrt{G} \mathrm{Im}(Z(\mathbf{X}_\theta)) \sqrt{G} \right| \right) \right\}$$

where $\mathbf{X}_\theta := (X_{\theta,1}, \ldots, X_{\theta,k})$ is a k-tuple of selfadjoint operators satisfying

$$\mathrm{Tr}(\rho_\theta X_{\theta,i}) = 0, \qquad \mathrm{Tr}\left(\frac{\partial \rho_\theta}{\partial \theta_i} X_{\theta,j}\right) = \delta_{i,j},$$

and $Z(\mathbf{X}_\theta)_{i,j} := (X_{\theta,i}, X_{\theta,j})_\theta = \mathrm{Tr}(\rho_\theta X_{\theta,j} X_{\theta,i})$.

The Holevo bound is achievable (asymptotically)

1. The Holevo bound is achieved in the case of quantum Gaussian shift models, i.e. Gaussian states of quantum cv systems with unknown means and fixed, known covariance.

2. The Holevo bound is achieved asymptotically for i.i.d. models of finite dimensional states, i.e. $\rho_\theta \otimes \cdots \otimes \rho_\theta$ with $\rho_\theta \in M(\mathbb{C}^d)$

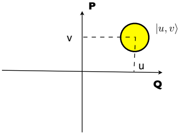
   The measurement consists of a two steps adaptive procedure (as in the case of one-dimensional parameter), with the difference that in the second step one needs to perform a joint measurement (not separable) on the $n - \tilde{n}$ systems. The measurement can be understood by showing that the n particle model 'converges' to a Gaussian model for which the solution is known.

   - A proof based on Cramér-Rao analysis is given for $d = 2$ in M. Hayashi and K. Matsumoto: arXiv:quant-ph/0411073
   - For the general case $d < \infty$ the result follows from the theory of 'local asymptotic normality' developed in J. Kahn and M. G. (CMP 2009)

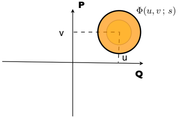
Quantum Gaussian shift model(s)

Displacement operator $D(u, v) := \exp(ivQ - iuP)$

- Coherent (laser) state
  $|u, v\rangle := D(u, v)|0\rangle$



- Displaced thermal state
  $\Phi(u, v\,; s) = D(u, v)\Phi(s)D(u, v)^*$



Estimation problem

Find the optimal measurement for $\{\Phi(u, v\,; s) : (u, v) \in \mathbb{R}^2\}$ with respect to

$$R_{max}(\hat{u}, \hat{v}) = \mathbb{E}[(u - \hat{u})^2 + (v - \hat{v})^2]$$

Joint measurement of Q and P

- Oscillator (Q, P) to be measured, prepared in state ρ

Joint measurement of Q and P

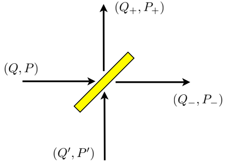
- Oscillator (Q, P) to be measured, prepared in state ρ
- Additional oscillator (Q', P') in state τ'

Joint measurement of Q and P

- Oscillator (Q, P) to be measured, prepared in state ρ
- Additional oscillator (Q', P') in state τ'

- Beam splitter

$$\begin{aligned} Q_\pm &:= Q \pm Q' \\ P_\pm &:= P \pm P' \end{aligned}$$



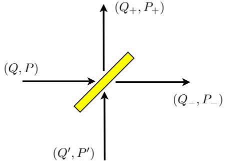
- Commuting noisy coordinates: $[Q_+, P_-] = 0$

Joint measurement of Q and P

- Oscillator (Q, P) to be measured, prepared in state ρ
- Additional oscillator (Q', P') in state τ'

- Beam splitter

$$Q_\pm := Q \pm Q'$$
$$P_\pm := P \pm P'$$



- Commuting noisy coordinates: $[Q_+, P_-] = 0$

Covariant measurements

Any displacement covariant measurement for (Q, P) is equivalent to measuring the pair $Q + \tilde{Q}$ and $P - \tilde{P}$ for some ancillary state τ of $(\tilde{Q}, \tilde{P})$.

Optimal measurement for Gaussian shift

- Gaussian shift model $\Phi(u, v; s) = D(u, v)\Phi(s)D^*(u, v)$
- Risk of the covariant measurement with (centred) ancilla state τ

$$R(\tau) = \mathrm{Tr}\left(\Phi(s) \otimes \tau \left((Q + \tilde{Q})^2 + (P - \tilde{P})^2\right)\right)$$
$$= \mathrm{Var}_{\Phi(s)}(Q) + \mathrm{Var}_{\Phi(s)}(P) + \mathrm{Var}_{\tau}(\tilde{Q}) + \mathrm{Var}_{\tau}(\tilde{P})$$

- Heterodyne measurement: when $\tau = |0\rangle\langle 0|$ the additional contribution is minimal

$$\mathrm{Var}_{|0\rangle\langle 0|}(\tilde{Q}) + \mathrm{Var}_{|0\rangle\langle 0|}(\tilde{P}) = 1$$

Theorem

The heterodyne measurement is optimal among covariant or unbiased measurements and achieves the minimax risk for the loss function $|u - \hat{u}|^2 + |v - \hat{v}|^2$.

5. Local asymptotic normality for i.i.d. quantum states

- The idea of local asymptotic normality
- Holstein-Primakov (Gaussian approximation)
- Local asymptotic normality for qubits
- Local asymptotic normality for d-dimensional systems

Reminder: local asymptotic normality for coin toss

- Data: $X_1, \ldots, X_n$ i.i.d. Bernoulli with $\mathbb{P}_\theta([X = 1]) = \theta$
- Optimal estimator: $\bar{X}_n = \sum_{i=1}^n X_i / n$
- Central Limit Theorem: $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1-\theta))$

  Local parameter: $\theta = \theta_0 + u/\sqrt{n}$

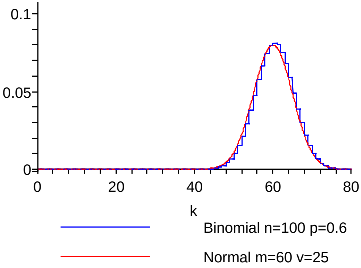
$$\hat{u}_n := \sqrt{n}(\hat{\theta}_n - \theta_0) \approx N(u, \theta_0(1-\theta_0))$$

Reminder: local asymptotic normality for coin toss

- Data: $X_1, \ldots, X_n$ i.i.d. Bernoulli with $\mathbb{P}_\theta([X = 1]) = \theta$
- Optimal estimator: $\bar{X}_n = \sum_{i=1}^n X_i / n$
- Central Limit Theorem: $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1-\theta))$

  Local parameter: $\theta = \theta_0 + u/\sqrt{n}$

$$\hat{u}_n := \sqrt{n}(\hat{\theta}_n - \theta_0) \approx N(u, \theta_0(1-\theta_0))$$

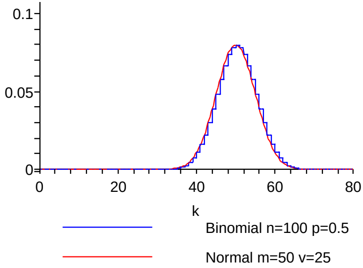


Reminder: local asymptotic normality for coin toss

- Data: $X_1, \ldots, X_n$ i.i.d. Bernoulli with $\mathbb{P}_\theta([X=1]) = \theta$
- Optimal estimator: $\bar{X}_n = \sum_{i=1}^n X_i / n$
- Central Limit Theorem: $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1-\theta))$

  Local parameter: $\theta = \theta_0 + u/\sqrt{n}$

$$\hat{u}_n := \sqrt{n}(\hat{\theta}_n - \theta_0) \approx N(u, \theta_0(1-\theta_0))$$

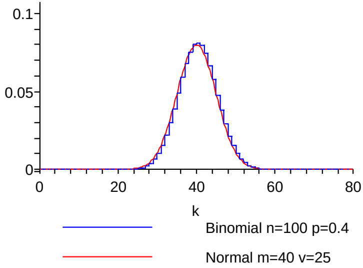


Reminder: local asymptotic normality for coin toss

- Data: $X_1, \ldots, X_n$ i.i.d. Bernoulli with $\mathbb{P}_\theta([X=1]) = \theta$
- Optimal estimator: $\bar{X}_n = \sum_{i=1}^n X_i / n$
- Central Limit Theorem: $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1-\theta))$

  Local parameter: $\theta = \theta_0 + u/\sqrt{n}$

$$\hat{u}_n := \sqrt{n}(\hat{\theta}_n - \theta_0) \approx N(u, \theta_0(1-\theta_0))$$

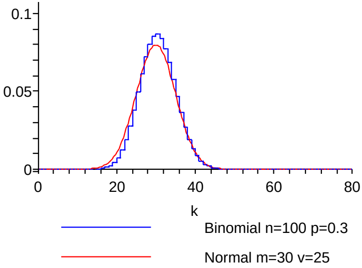


Reminder: local asymptotic normality for coin toss

- Data: $X_1, \ldots, X_n$ i.i.d. Bernoulli with $\mathbb{P}_\theta([X=1]) = \theta$
- Optimal estimator: $\bar{X}_n = \sum_{i=1}^n X_i/n$
- Central Limit Theorem: $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1-\theta))$

Local parameter: $\theta = \theta_0 + u/\sqrt{n}$

$$\hat{u}_n := \sqrt{n}(\hat{\theta}_n - \theta_0) \approx N(u, \theta_0(1-\theta_0))$$



LAN for general parametric model

- $(Y_1, \ldots, Y_n)$ i.i.d. with $\mathbb{P}^{\theta_0 + u/\sqrt{n}}$ a 'smooth' family with $u \in \mathbb{R}^k$. Then

$$\left\{ \mathbb{P}^n_{\theta_0 + u/\sqrt{n}} : u \in \mathbb{R}^k \right\} \rightsquigarrow \left\{ N(u, I_{\theta_0}^{-1}) : u \in \mathbb{R}^k \right\}$$

LAN for general parametric model

- $(Y_1, \dots, Y_n)$ i.i.d. with $\mathbb{P}^{\theta_0+u/\sqrt{n}}$ a 'smooth' family with $u \in \mathbb{R}^k$. Then

$$\left\{\mathbb{P}^n_{\theta_0+u/\sqrt{n}} : u \in \mathbb{R}^k\right\} \rightsquigarrow \left\{N(u, I_{\theta_0}^{-1}) : u \in \mathbb{R}^k\right\}$$

- Weak convergence:

$$\left\{\frac{\mathrm{d}\mathbb{P}^n_{\theta_0+u/\sqrt{n}}}{\mathrm{d}\mathbb{P}^n_{\theta_0}} : u \in \mathbb{R}^k\right\} \xrightarrow{\mathcal{D}} \left\{\frac{\mathrm{d}N(u, I_{\theta_0}^{-1})}{\mathrm{d}N(0, I_{\theta_0}^{-1})} : u \in \mathbb{R}^k\right\}$$

LAN for general parametric model

- $(Y_1, \ldots, Y_n)$ i.i.d. with $\mathbb{P}^{\theta_0+u/\sqrt{n}}$ a 'smooth' family with $u \in \mathbb{R}^k$. Then

$$\left\{\mathbb{P}^n_{\theta_0+u/\sqrt{n}} : u \in \mathbb{R}^k\right\} \rightsquigarrow \left\{N(u, I_{\theta_0}^{-1}) : u \in \mathbb{R}^k\right\}$$

- Strong convergence (Le Cam):

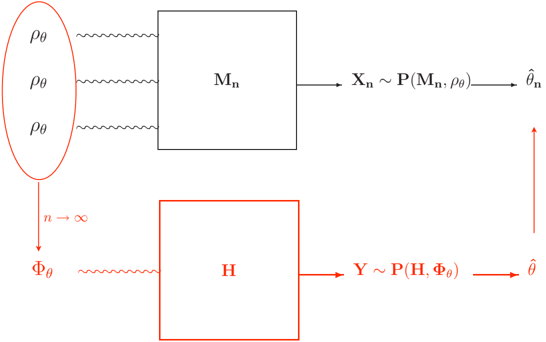
  there exist randomizations T_n, S_n such that for $\eta < 1/4$

$$\lim_{n\to\infty} \sup_{\|u\|\le n^\eta} \left\| T_n \, \mathbb{P}^n_{\theta_0+u/\sqrt{n}} - N(u, I_{\theta_0}^{-1}) \right\|_{\mathrm{tv}} = 0$$

and

$$\lim_{n\to\infty} \sup_{\|u\|\le n^\eta} \left\| \mathbb{P}^n_{\theta_0+u/\sqrt{n}} - S_n \, N(u, I_{\theta_0}^{-1}) \right\|_{\mathrm{tv}} = 0$$

Optimal estimation using local asymptotic normality



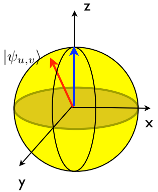
[L. Le Cam]

- Sequence of I.I.D. quantum statistical models $\mathcal{Q}_n = \{\rho_\theta^{\otimes n} : \theta \in \Theta\}$
- $\mathcal{Q}_n$ converges (locally) to simpler Gaussian shift model $\mathcal{Q}$
- Optimal measurement for limit $\mathcal{Q}$ can be pulled back to $\mathcal{Q}_n$

Two quantum state estimation problems

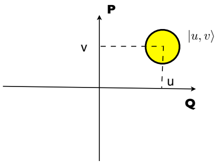
- Two parameter model in $\mathbb{C}^2$

  $|\psi_{u,v}\rangle = \exp(i(v\sigma_x - u\sigma_y))|\uparrow\rangle$



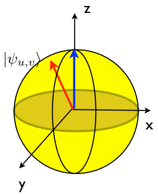
- Coherent (laser) state

  $|u, v\rangle = D(u, v)|0\rangle$



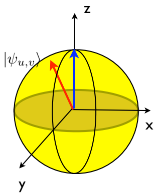
Estimation of a pure spin state revisited

Two-dim. model: (small) rotation of $|\uparrow\rangle$



$$|\psi_{u,v}\rangle := \exp\left(i(u\sigma_x - v\sigma_y)\right)|\uparrow\rangle$$

Estimation of a pure spin state revisited



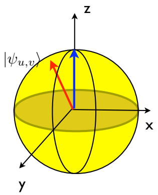
Two-dim. model: (small) rotation of $|\uparrow\rangle$

$$|\psi_{u,v}\rangle := \exp\left(i(u\sigma_x - v\sigma_y)\right)|\uparrow\rangle$$

Symmetric logarithmic derivatives at $(u, v) = (0, 0)$:

$$\begin{cases} \left.\frac{\partial \rho_{u,v}}{\partial u}\right|_{u=0,v=0} = \rho_{0,0} \circ \mathcal{L}_{0,0}^{(u)} & \Longrightarrow \mathcal{L}_{0,0}^{(u)} = 2\sigma_y \\ \left.\frac{\partial \rho_{u,v}}{\partial u}\right|_{u=0,v=0} = \rho_{0,0} \circ \mathcal{L}_{0,0}^{(v)} & \Longrightarrow \mathcal{L}_{0,0}^{(v)} = 2\sigma_x \end{cases}$$

Estimation of a pure spin state revisited



Two-dim. model: (small) rotation of $|\uparrow\rangle$

$$|\psi_{u,v}\rangle := \exp\left(i(u\sigma_x - v\sigma_y)\right)|\uparrow\rangle$$

Symmetric logarithmic derivatives at $(u,v) = (0,0)$:

$$\begin{cases} \left.\frac{\partial \rho_{u,v}}{\partial u}\right|_{u=0,v=0} = \rho_{0,0} \circ \mathcal{L}^{(u)}_{0,0} \quad \Longrightarrow \mathcal{L}^{(u)}_{0,0} = 2\sigma_y \\ \left.\frac{\partial \rho_{u,v}}{\partial u}\right|_{u=0,v=0} = \rho_{0,0} \circ \mathcal{L}^{(v)}_{0,0} \quad \Longrightarrow \mathcal{L}^{(v)}_{0,0} = 2\sigma_x \end{cases}$$

Expectations and variances:

$$\begin{cases} \mathbb{E}[2\sigma_y] \approx 4u \qquad \mathrm{Var}(2\sigma_y) = 4\mathrm{Tr}\left[\rho_{u,v}(\sigma_y - \mathbb{E}[\sigma_y]\mathbf{1})^2\right] \approx 4 \\ \mathbb{E}[2\sigma_x] \approx 4v \qquad \mathrm{Var}(2\sigma_x) = 4\mathrm{Tr}\left[\rho_{u,v}(\sigma_x - \mathbb{E}[\sigma_x]\mathbf{1})^2\right] \approx 4 \end{cases}$$

Optimal measurements for u_x and u_y are incompatible: $[\sigma_x, \sigma_y] \neq 0$

Holstein-Primakoff (Gaussian approximation)

- n identically prepared spin-1/2 systems

$$\left|\psi_{\frac{u}{\sqrt{n}},\frac{v}{\sqrt{n}}}\right\rangle := \exp\left(i\frac{v\sigma_x - u\sigma_y}{\sqrt{n}}\right)|\uparrow\rangle$$

- Collective observables $L_{x,y,z} := \sum_{i=1}^{n} \sigma_{x,y,z}^{(i)}$

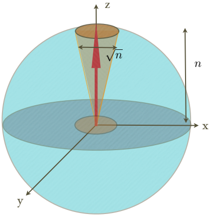
- Quantum Central Limit Theorem $(u = 0, v = 0)$

$$\frac{L_x}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1)$$

$$\frac{L_y}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1)$$

$$\left[\frac{L_x}{\sqrt{n}}, \frac{L_y}{\sqrt{n}}\right] = \frac{2i}{n}L_z \xrightarrow{l.l.n.} 2i\mathbf{1}$$

$$\left[\frac{L_y}{\sqrt{n}}, \frac{L_z}{\sqrt{n}}\right] = \frac{2i}{n}L_x \xrightarrow{l.l.n.} 0$$



[Holstein and Primakoff P.R. 1940] [Radcliffe J.Phys. A 1971] [Klein and Marshalek Rev. Mod. Phys. 1991]

Holstein-Primakoff (Gaussian approximation)

- n identically prepared spin-1/2 systems

$$\left|\psi_{\frac{u}{\sqrt{n}},\frac{v}{\sqrt{n}}}\right\rangle := \exp\left(i\frac{v\sigma_x - u\sigma_y}{\sqrt{n}}\right)|\uparrow\rangle$$

- Collective observables $L_{x,y,z} := \sum_{i=1}^{n} \sigma_{x,y,z}^{(i)}$

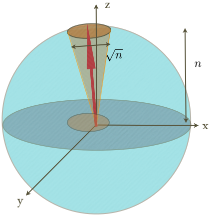
- Quantum Central Limit Theorem $(u \neq 0, v \neq 0)$

$$\frac{L_x}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(2u, 1)$$

$$\frac{L_y}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(2v, 1)$$

$$\left[\frac{L_x}{\sqrt{n}}, \frac{L_y}{\sqrt{n}}\right] = \frac{2i}{n} L_z \xrightarrow{l.l.n.} 2i\mathbf{1}$$

$$\left[\frac{L_y}{\sqrt{n}}, \frac{L_z}{\sqrt{n}}\right] = \frac{2i}{n} L_x \xrightarrow{l.l.n.} 0$$



[Holstein and Primakoff P.R. 1940] [Radcliffe J.Phys. A 1971] [Klein and Marshalek Rev. Mod. Phys. 1991]

Holstein-Primakoff (Gaussian approximation)

- n identically prepared spin-1/2 systems

$$\left|\psi_{\frac{u}{\sqrt{n}},\frac{v}{\sqrt{n}}}\right\rangle := \exp\left(i\frac{v\sigma_x - u\sigma_y}{\sqrt{n}}\right)|\uparrow\rangle$$

- Collective observables $L_{x,y,z} := \sum_{i=1}^{n} \sigma_{x,y,z}^{(i)}$

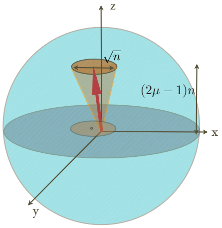
- Quantum Central Limit Theorem (mixed states)

$$\frac{L_x}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(2(2\mu-1)u, 1)$$

$$\frac{L_z - n(2\mu-1)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(h, \mu(1-\mu))$$

$$\left[\frac{L_x}{\sqrt{n}}, \frac{L_y}{\sqrt{n}}\right] = \frac{2i}{n}L_z \xrightarrow{l.l.n.} 2(2\mu-1)i\mathbf{1}$$

$$\left[\frac{L_y}{\sqrt{n}}, \frac{L_z}{\sqrt{n}}\right] = \frac{2i}{n}L_x \xrightarrow{l.l.n.} 0$$

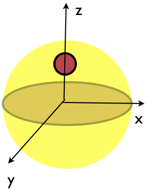


[Holstein and Primakoff P.R. 1940] [Radcliffe J.Phys. A 1971] [Klein and Marshalek Rev. Mod. Phys. 1991]

Local spin model and the Gaussian limit

- $\left\{\rho_{\mathbf{u}/\sqrt{n}} \ : \ \mathbf{u} = (u, v, h)\right\}$ neighbourhood of $\rho_0 := \mathrm{Diag}(\mu, 1-\mu)$

$$\rho_{\mathbf{u}/\sqrt{n}} := U_n(u,v) \begin{bmatrix} \mu + \frac{h}{\sqrt{n}} & 0 \\ 0 & 1 - \mu - \frac{h}{\sqrt{n}} \end{bmatrix} U_n(u,v)^*$$

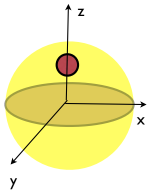


$$U_n(u,v) := \exp(i(v\sigma_x - u\sigma_y)/\sqrt{n})$$

- $\{\rho_{\mathbf{u}/\sqrt{n}} \ : \ \mathbf{u} = (u, v, h)\}$ neighbourhood of $\rho_0 := \mathrm{Diag}(\mu, 1-\mu)$

$$\rho_{\mathbf{u}/\sqrt{n}} := U_n(u,v) \begin{bmatrix} \mu + \frac{h}{\sqrt{n}} & 0 \\ 0 & 1-\mu-\frac{h}{\sqrt{n}} \end{bmatrix} U_n(u,v)^*$$

$$U_n(u,v) := \exp(i(v\sigma_x - u\sigma_y)/\sqrt{n})$$



- Gaussian shift model: $N_{\mathbf{u}} \otimes \Phi_{\mathbf{u}}$
  - Classical part: $\frac{L_z - n(2\mu-1)}{\sqrt{n}} \longrightarrow X$ with distribution $N_{\mathbf{u}} := N(h, \mu(1-\mu))$
  - Quantum part:

$$\left.\begin{array}{l} \frac{L_x}{\sqrt{2n(2\mu-1)}} \longrightarrow Q \\ \frac{L_y}{\sqrt{2n(2\mu-1)}} \longrightarrow P \end{array}\right\} \text{in state } \Phi_{\mathbf{u}} := \Phi\left(u\sqrt{2(2\mu-1)}, \, v\sqrt{2(2\mu-1)} \, ; \, \frac{1}{2(2\mu -\right.$$

Theorem

Let $\rho_{\mathbf{u},n} := \left(\rho_{\mathbf{u}/\sqrt{n}}\right)^{\otimes n}$ be the state of n i.i.d. spins with $1/2 < \mu < 1$.

Then there exist quantum channels T_n, S_n such that for any $\eta < 1/4$

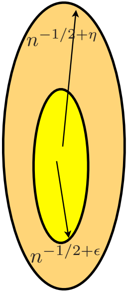
$$\lim_{n\to\infty} \sup_{\|\mathbf{u}\|<n^{\eta}} \| T_n\left(\rho_{\mathbf{u},n}\right) - N_{\mathbf{u}} \otimes \Phi_{\mathbf{u}}\|_1 = 0,$$

and

$$\lim_{n\to\infty} \sup_{\|\mathbf{u}\|<n^{\eta}} \|\rho_{\mathbf{u},n} - S_n\left(N_{\mathbf{u}} \otimes \Phi_{\mathbf{u}}\right)\|_1 = 0.$$

[Guta, Janssens and Kahn, C.M.P. 2008]

Asymptotically optimal (adaptive) measurement procedure



Given n i.i.d. spins prepared in state ρ_θ

1. Use $n^{1-\epsilon}$ copies to produce a rough estimator ρ_0
2. Map remaining $\tilde{n} = n - n^{1-\epsilon}$ states through $T_{\tilde{n}}$
3. Perform optimal Gaussian measurement and produce estimator

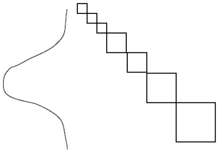
$$\hat{\theta}_n = \theta_0 + \hat{\mathbf{u}}/\sqrt{\tilde{n}}$$

Idea of the proof

- Block diagonal form (Weyl Theorem)

$$\left(\mathbb{C}^2\right)^{\otimes n} = \bigoplus_{j=0,1/2}^{n/2} \mathbb{C}^{2j+1} \otimes \mathbb{C}^{d_j}$$

$$\rho_{\mathbf{u}/\sqrt{n}}^{\otimes n} = \bigoplus_{j=0,1/2}^{n/2} p_{\mathbf{u},n}(j)\, \rho_{\mathbf{u},n}(j) \otimes \frac{\mathbf{1}}{d_j}$$



- Classical part: $p_{\mathbf{u},n}(j) = \mathbb{P}[L = j]$ with L the total spin

$$L \approx L_z \sim \mathrm{Bin}(\mu + u_z/\sqrt{n}, n) \xrightarrow{s.} N_{\mathbf{u}}$$

- Quantum part: embed conditional state $\rho_{\mathbf{u},j}$ isometrically into $L^2(\mathbb{R})$

$$\begin{aligned} V_j &: \mathcal{H}_j \to L^2(\mathbb{R}) \\ T_j &: \rho_{\mathbf{u},j} \longmapsto V_j \rho_{\mathbf{u},j} V_j^* \end{aligned}$$

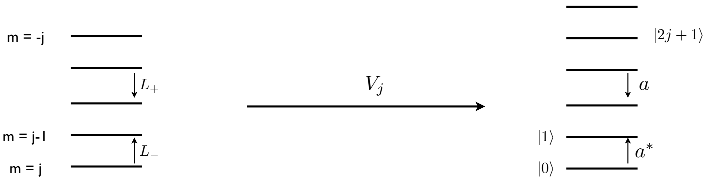
Isometric embedding

- Orthonormal bases

$$
\begin{aligned}
L_z|m,j\rangle &= m|m,j\rangle && (\ \mathbb{C}^{2j+1}\ ) \\
|k\rangle &= H_k(x)e^{-x^2/2} && (\ L^2(\mathbb{R})\ )
\end{aligned}
$$

- Ladder operators

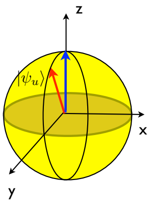
$$
\begin{cases} L_+ & := L_x + iL_y \\ L_- & := L_x - iL_y \end{cases} \quad \text{and} \quad \begin{cases} a & := (Q+iP)/\sqrt{2} \\ a^* & := (Q-iP)/\sqrt{2} \end{cases}
$$



Back to pure spin states

- *n* identically prepared spins

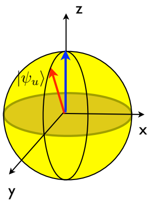
$$|\psi_{u,n}\rangle = \left[\cos\left(\frac{u}{\sqrt{n}}\right) |\uparrow\rangle + \sin\left(\frac{u}{\sqrt{n}}\right) |\downarrow\rangle\right]^{\otimes n}$$



Back to pure spin states

- n identically prepared spins

$$|\psi_{u,n}\rangle = \left[\cos\left(\frac{u}{\sqrt{n}}\right) |\uparrow\rangle + \sin\left(\frac{u}{\sqrt{n}}\right) |\downarrow\rangle\right]^{\otimes n}$$



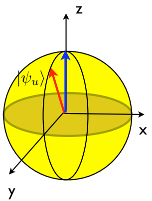
Local asymptotic normality

$\{|\psi_{u,n}\rangle : u \in \mathbb{R}\}$ converges to the Gaussian model $\{|\sqrt{2}u, 0\rangle : u \in \mathbb{R}\}$

Back to pure spin states

- n identically prepared spins

$$|\psi_{u,n}\rangle = \left[\cos\left(\frac{u}{\sqrt{n}}\right)|\uparrow\rangle + \sin\left(\frac{u}{\sqrt{n}}\right)|\downarrow\rangle\right]^{\otimes n}$$



Local asymptotic normality

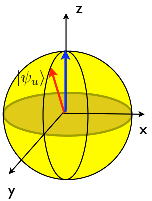
$\{|\psi_{u,n}\rangle : u \in \mathbb{R}\}$ converges to the Gaussian model $\{|\sqrt{2}u, 0\rangle : u \in \mathbb{R}\}$

- Weak convergence:

$$\langle\psi_{u,n}|\psi_{v,n}\rangle = \cos((u-v)/\sqrt{n})^n \longrightarrow e^{-\frac{1}{2}(u-v)^2} = \langle\sqrt{2}u, 0|\sqrt{2}v, 0\rangle$$

- n identically prepared spins

$$|\psi_{u,n}\rangle = \left[\cos\left(\frac{u}{\sqrt{n}}\right) |\uparrow\rangle + \sin\left(\frac{u}{\sqrt{n}}\right) |\downarrow\rangle\right]^{\otimes n}$$



Local asymptotic normality

$\{|\psi_{u,n}\rangle : u \in \mathbb{R}\}$ converges to the Gaussian model $\{|\sqrt{2}u, 0\rangle : u \in \mathbb{R}\}$

- Weak convergence:

$$\langle\psi_{u,n}|\psi_{v,n}\rangle = \cos((u-v)/\sqrt{n})^n \longrightarrow e^{-\frac{1}{2}(u-v)^2} = \langle\sqrt{2}u, 0|\sqrt{2}v, 0\rangle$$

- Strong convergence: there exist quantum channels T_n s.t. for $0 < \eta < 1/4$

$$\lim_{n\to\infty} \sup_{\|\mathbf{u}\|\leq n^{\eta}} \left\| T_n\left(|\psi_{u,n}\rangle\langle\psi_{u,n}|\right) - |\sqrt{2}u, 0\rangle\langle\sqrt{2}u, 0| \right\|_1 = 0$$

Local asymptotic normality in d-dimensions

- Local model around $\rho_0 = \mathrm{Diag}(\mu_1, \ldots, \mu_d)$ with $\mu_1 > \mu_2 > \cdots > \mu_d > 0$

$$\rho_{\mathbf{u}/\sqrt{n}} = \begin{bmatrix} \mu_1 + h_1/\sqrt{n} & \ldots & z^*_{1,d}/\sqrt{n} \\ \vdots & \ddots & \vdots \\ z_{1,d}/\sqrt{n} & \ldots & \mu_d - \sum_{i=1}^{d-1} h_i/\sqrt{n} \end{bmatrix} \quad \mathbf{u} = (\mathbf{h}, \mathbf{z}) \in \mathbb{R}^{d-1} \times \mathbb{C}^{d(d-1)/2}$$

- Gaussian shift model: $N_{\mathbf{u}} \otimes \Phi_{\mathbf{u}}$
  - Classical part: $N_{\mathbf{u}} := N(\mathbf{h}, I_\mu^{-1})$
  - Quantum part: $\Phi_{\mathbf{u}} := \bigotimes_{1 \le j < k \le d} \Phi\left(\frac{z_{j,k}}{2\sqrt{\mu_j - \mu_k}} ; \frac{\mu_j + \mu_k}{2(\mu_j - \mu_k)} \right)$

Local asymptotic normality in d-dimensions

Theorem

Let $\rho_{\mathbf{u},n} := \left(\rho_{\mathbf{u}/\sqrt{n}}\right)^{\otimes n}$ be the state of n i.i.d systems with $\mu_1 > \cdots > \mu_d > 0$.

Then there exist quantum channels T_n, S_n such that

$$\lim_{n\to\infty} \sup_{\mathbf{u}\in\Theta_{n,\beta,\gamma}} \left\| T_n(\rho_{\mathbf{u},n}) - N_{\mathbf{u}} \otimes \Phi_{\mathbf{u}} \right\|_1 = 0$$

$$\lim_{n\to\infty} \sup_{\mathbf{u}\in\Theta_{n,\beta,\gamma}} \left\| S_n(N_{\mathbf{u}} \otimes \Phi_{\mathbf{u}}) - \rho_{\mathbf{u},n} \right\|_1 = 0$$

where

$$\Theta_{n,\beta,\gamma} = \left\{ \mathbf{u} := (\mathbf{h},\mathbf{d}) : \|\mathbf{z}\| \leq n^{\beta}, \|\mathbf{d}\| \leq n^{\gamma} \right\}, \text{ with } \beta < 1/9, \gamma < 1/4.$$

[Guta and Kahn, C.M.P. 2009]

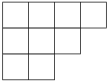
Blocks indexed by Young diagrams

- Block diagonal form

$$\left(\mathbb{C}^d\right)^{\otimes n} = \bigoplus_\lambda \mathcal{H}_\lambda \otimes \mathcal{K}_\lambda$$

$$\rho_{\mathbf{u}/\sqrt{n}}^{\otimes n} = \bigoplus_\lambda p_{\mathbf{u},n}(\lambda)\, \rho_{\mathbf{u},n}(\lambda) \otimes \mathrm{tr}_\lambda$$

- Young diagrams λ with d lines and n boxes



- Classical part: $p_{\mathbf{u},n} \approx \mathrm{Mult}\left(\mu_1 + \frac{h_1}{\sqrt{n}}, \ldots, \mu_d - \sum_i \frac{h_i}{\sqrt{n}}; n\right) \Longrightarrow N_{\mathbf{u}}$

Bases and ladder operators in $\mathcal{H}_\lambda$

- Non-orthogonal basis $|t, \lambda\rangle = |\mathbf{m}, \lambda\rangle$

  $\mathbf{m} = (m_{i,j} = \sharp\text{j's in row i}\} : i < j)$

$$\begin{array}{|c|c|c|}\hline 1&1&2\\\hline 2&2\\\cline{1-2} 3\\\cline{1-1}\end{array}$$

semi-standard Young tableau t

- Typical vectors are $\approx$ orthogonal

  If $|\mathbf{m}|, |\mathbf{l}| = O(n^\eta)$ with $\eta < 2/9$ then

$$|\langle \mathbf{m}, \lambda \,|\, \mathbf{l}, \lambda\rangle| = O(n^{-c(\eta)})$$

$$\begin{array}{|c|c|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1&2&2&3\\\hline 2&2&2&2&3&3\\\cline{1-6} 3&3&3\\\cline{1-3}\end{array}$$

typical Young tableau t

  - Approximate ladder operators

$$L^*_{2,3} : \begin{array}{|c|c|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1&2&2&3\\\hline 2&2&2&2&3&3\\\cline{1-6} 3&3&3\\\cline{1-3}\end{array} \longrightarrow O(n^\eta) \begin{array}{|c|c|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1&\boxed{2}&3&3\\\hline 2&2&2&2&3&3\\\cline{1-6} 3&3&3\\\cline{1-3}\end{array} + O(n) \begin{array}{|c|c|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1&2&2&3\\\hline 2&2&2&\boxed{3}&3&3\\\cline{1-6} 3&3&3\\\cline{1-3}\end{array}$$

  - Approximate isometry

$$V_\lambda : |\mathbf{m}\rangle \longmapsto \bigotimes_{1\le j<k\le d} |m_{j,k}\rangle$$

Application of LAN: asymptotically optimal learning of qubit states

- Helstrom measurement: optimal discrimination between 2 known states ρ_0 and ρ_1 with prior probabilities (π_0, π_1)

$$M_0 := [\pi_0 \rho_0 - \pi_1 \rho_1]_+ \quad M_1 := [\pi_0 \rho_0 - \pi_1 \rho_1]_-$$

- Optimal error:

$$\pi_0 \mathrm{Tr}(\rho_0 M_1) + \pi_1 \mathrm{Tr}(\rho_1 M_0) = \frac{1}{2}(1 - \mathrm{Tr}(|\pi_0 \rho_0 - \pi_1 \rho_1|))$$

Application of LAN: asymptotically optimal learning of qubit states

- Helstrom measurement: optimal discrimination between 2 known states ρ_0 and ρ_1 with prior probabilities (π_0, π_1)

$$M_0 := [\pi_0 \rho_0 - \pi_1 \rho_1]_+ \quad M_1 := [\pi_0 \rho_0 - \pi_1 \rho_1]_-$$

- Optimal error:

$$\pi_0 \mathrm{Tr}(\rho_0 M_1) + \pi_1 \mathrm{Tr}(\rho_1 M_0) = \frac{1}{2}(1 - \mathrm{Tr}(|\pi_0 \rho_0 - \pi_1 \rho_1|))$$

- Qubit learning:
  - given n labelled qubits drawn from unknown ρ_0 and ρ_1 with probabilities (π_0, π_1)

Application of LAN: asymptotically optimal learning of qubit states

- Helstrom measurement: optimal discrimination between 2 known states ρ_0 and ρ_1 with prior probabilities (π_0, π_1)

$$M_0 := [\pi_0\rho_0 - \pi_1\rho_1]_+ \quad M_1 := [\pi_0\rho_0 - \pi_1\rho_1]_-$$

- Optimal error:

$$\pi_0 \mathrm{Tr}(\rho_0 M_1) + \pi_1 \mathrm{Tr}(\rho_1 M_0) = \frac{1}{2}(1 - \mathrm{Tr}(|\pi_0\rho_0 - \pi_1\rho_1|))$$

- Qubit learning:
  - given n labelled qubits drawn from unknown ρ_0 and ρ_1 with probabilities (π_0, π_1)
  - task: learn the optimal measurement M to be used for future state discrimination

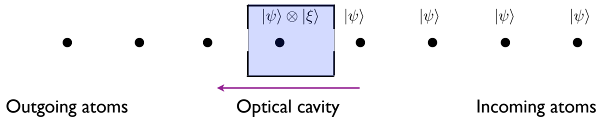
- Using LAN it can be shown that the optimal solution is not the plug-in estimator of M based on optimal state estimation, but a joint measurement of the training set $\rho_0^{\otimes n_0} \otimes \rho_1^{\otimes n_1}$

[M. Guta, W. Kotlowski, N.J.P. 2010]

System identification for quantum Markov processes

- Quantum Markov chains
- Mixing chains
- LAN and quantum Fisher information of the output state
- LAN for simple measurements on the output

Quantum Markov chains



- Examples: quantum optical networks, atom maser, solid state cavity QED...

- Dynamics: unitary 'scattering' of atoms by cavity

$$U : M(\mathbb{C}^d \otimes \mathbb{C}^k) \to M(\mathbb{C}^d \otimes \mathbb{C}^k)$$

- Discrete time version of quantum Markov processes driven by white noise

- Closely related to Matrix Product States (MPS) and Channels with Memory

[Kümmerer, J.F.A. 1985] [Fannes, Nachtergale and Werner C.M.P. 1992] [Kretschmann and Werner, P.R. A 2005]

Examples

- Jaynes-Cummings coupling

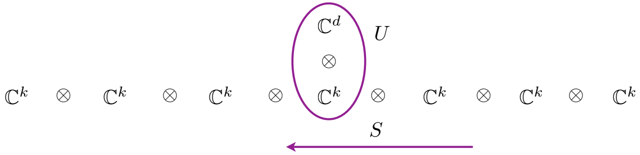
$$
\begin{aligned}
U &: \mathbb{C}^2 \otimes \ell^2(\mathbb{N}) \to \mathbb{C}^2 \otimes \ell^2(\mathbb{N}) \\
U &= \exp\left[\alpha(\sigma_- \otimes a^* + \sigma_+ \otimes a) + i\beta\sigma_z + i\gamma a^* a\right]
\end{aligned}
$$

- Continuous-time quantum Markov process

$$
\begin{aligned}
U_t &: \mathbb{C}^d \otimes \mathcal{F}(L^2(\mathbb{R}_+)) \to \mathbb{C}^d \otimes \mathcal{F}(L^2(\mathbb{R}_+)) \\
dU_t &= \left\{ L \otimes dA_t^* - L^* \otimes dA_t - \frac{1}{2} L^* L\, dt - iH\, dt \right\} U_t \qquad \text{(QSDE)}
\end{aligned}
$$

Hilbert space evolution

- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U



- One step joint evolution: $W = S \circ U$

Hilbert space evolution

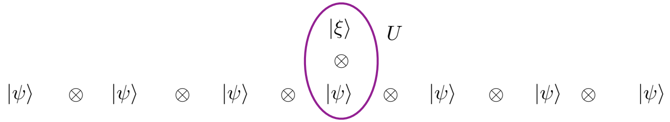
- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U

$$\begin{matrix} & & & & & & |\xi\rangle & & & & & & \\ & & & & & & \otimes & & & & & & \\ |\psi\rangle & \otimes & |\psi\rangle & \otimes & |\psi\rangle & \otimes & |\psi\rangle & \otimes & |\psi\rangle & \otimes & |\psi\rangle & \otimes & |\psi\rangle \end{matrix}$$

- One step joint evolution: $W = S \circ U$

Hilbert space evolution

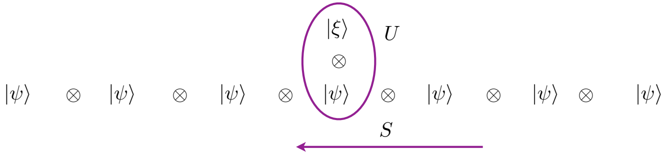
- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U



- One step joint evolution: $W = S \circ U$

Hilbert space evolution

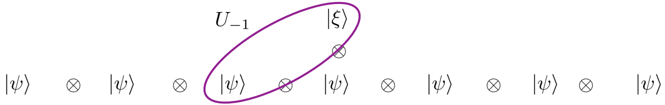
- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U



- One step joint evolution: $W = S \circ U$

Hilbert space evolution

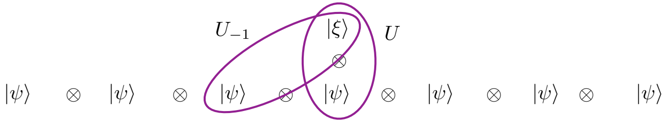
- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U



- One step joint evolution: $W = S \circ U$

Hilbert space evolution

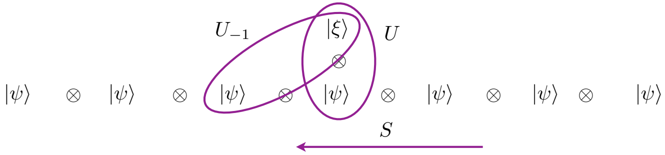
- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U



- One step joint evolution: $W = S \circ U$

Hilbert space evolution

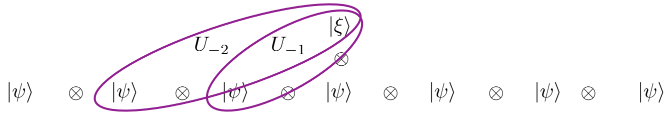
- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U



- One step joint evolution: $W = S \circ U$

Hilbert space evolution

- 'system' $\mathbb{C}^d$, 'noise unit' $\mathbb{C}^k$, interaction unitary U



- One step joint evolution: $W = S \circ U$
- Output state after n steps

$$|\psi_n\rangle := U_{-1} \circ \cdots \circ U_{-n} \, |\xi\rangle \otimes |\psi\rangle^{\otimes n} \in \mathbb{C}^d \otimes \left(\mathbb{C}^k\right)^{\otimes n}$$

Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

$$X \mapsto T(X) := \left\langle \psi \right| U^{-1} (X \otimes \mathbf{1}) \, U \left| \psi \right\rangle$$

$$\begin{array}{ccccccccccccc} & & & & & & X & & & & & & \\ & & & & & & \otimes & & & & & & \\ \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} \end{array}$$

Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

$$X \mapsto T(X) := \left\langle \psi \,\middle|\, U^{-1} \, (X \otimes \mathbf{1}) \, U \,\middle|\, \psi \right\rangle$$

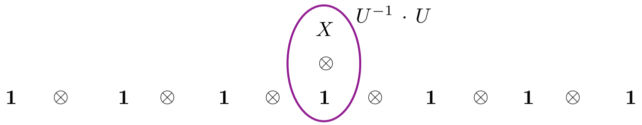
$$\begin{matrix} & & & & & & X & & & & & & \\ & & & & & & \otimes & & & & & & \\ \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} \end{matrix}$$

$$\xrightarrow{S^{-1} \cdot S}$$

Markov (transition) semigroup

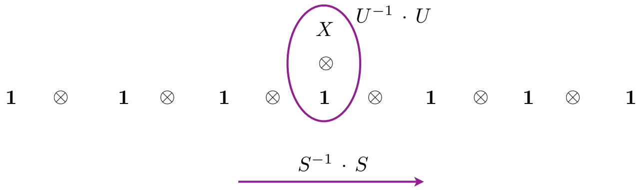
- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

$$X \mapsto T(X) := \left\langle \psi \right| U^{-1} (X \otimes \mathbf{1}) \, U \left| \psi \right\rangle$$



- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

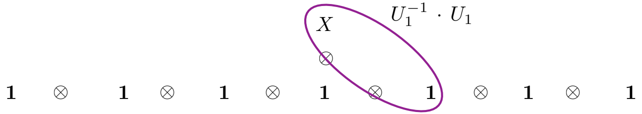
$$X \mapsto T(X) := \left\langle \psi \right| U^{-1} \, (X \otimes \mathbf{1}) \; U \left| \psi \right\rangle$$



Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

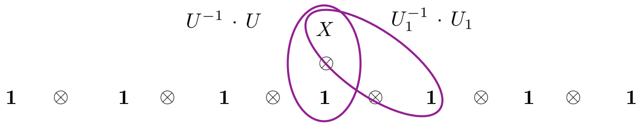
$$X \mapsto T(X) := \left\langle \psi \right| U^{-1} \left(X \otimes \mathbf{1} \right) U \left| \psi \right\rangle$$



Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

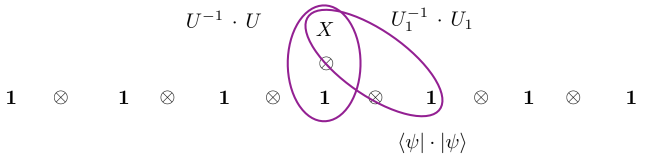
$$X \mapsto T(X) := \left\langle \psi \right| U^{-1} \, (X \otimes \mathbf{1}) \; U \left| \psi \right\rangle$$



Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

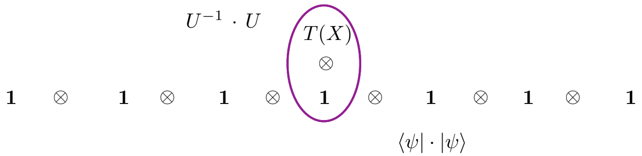
$$X \mapsto T(X) := \left\langle \psi \right| U^{-1} \left(X \otimes \mathbf{1}\right) U \left| \psi \right\rangle$$



Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

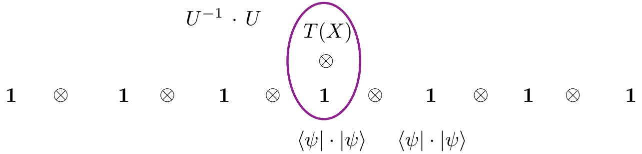
$$X \mapsto T(X) := \left\langle \psi \,\middle|\, U^{-1}\,(X \otimes \mathbf{1})\; U \,\middle|\, \psi \right\rangle$$



Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

$$X \mapsto T(X) := \left\langle \psi \,\middle|\, U^{-1}\,(X \otimes \mathbf{1})\, U \,\middle|\, \psi \right\rangle$$



Markov (transition) semigroup

- $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$ describes the 'reduced' evolution of the system

$$X \mapsto T(X) := \left\langle \psi \,\middle|\, U^{-1} \, (X \otimes \mathbf{1}) \, U \,\middle|\, \psi \right\rangle$$

$$\begin{array}{ccccccccccccc} & & & & & & T^2(X) & & & & & & \\ & & & & & & \otimes & & & & & & \\ \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} & \otimes & \mathbf{1} \\ & & & & & & \langle\psi|\cdot|\psi\rangle & & \langle\psi|\cdot|\psi\rangle & & & & \end{array}$$

- after n steps

$$X \mapsto T_n(X) := \left\langle \psi^{\otimes n} \,\middle|\, \hat{U}^{-n} \, (X \otimes \mathbf{1}) \, \hat{U}^{n} \,\middle|\, \psi^{\otimes n} \right\rangle = T^n(X)$$

Mixing (ergodic) quantum Markov chain

- Transition operator $T : M(\mathbb{C}^d) \to M(\mathbb{C}^d)$

$$T(X) := \left\langle \psi \right| U^\dagger \, (X \otimes \mathbf{1}) \, U \left| \psi \right\rangle$$

- Mixing Markov chain (transition operator T)
  - $T(X) = X$ if and only if $X = \alpha \mathbf{1}$
  - All other eigenvalues λ satisfy $|\lambda| < 1$.

- Convergence to equilibrium

  If T is mixing then there exists a unique stationary state ρ_∞ on $M(\mathbb{C}^d)$ and

$$\lim_{n \to \infty} T_*^n(\sigma) = \rho_\infty, \qquad \text{for all initial states } \sigma$$

- Classical analogue

  Finite state irreducible aperiodic chain (Perron-Frobenius Therem)

L.A.N. for (one parameter) coupling constant

- Let $U_\theta = \exp(i\theta K)$ with unknown θ, and assume that T is mixing.
- Let $|\psi_{u,n}\rangle$ be the output state (statistical model)

$$|\psi_{u,n}\rangle := \left(S \circ U_{\theta_0+u/\sqrt{n}}\right)^n \left|\xi \otimes \psi^{\otimes n}\right\rangle$$

Theorem

1. *the quantum Fisher information scales (asymptotically) linearly*

$$\frac{1}{n} H_n(\theta_0) \to H$$

2. $|\psi_{u,n}\rangle$ *is asymptotically normal, i.e*

$$\lim_{n\to\infty} \langle \psi_{u,n} | \psi_{v,n} \rangle = \langle \sqrt{H/2}u \,|\, \sqrt{H/2}v \rangle$$

where $\{|\sqrt{H/2}u\rangle : u \in \mathbb{R}\}$ *is the quantum Gaussian shift with Fisher info* H.

Fisher information = variance of generator

- The asymptotic Fisher information is $H(\theta_0) = 4V(K,K)$ with 'variance'

$$V(K,K) \quad := \quad \mathbb{E}\left(K^2\right) + 2\mathbb{E}\left(K \circ (\mathrm{Id} - T_{\theta_0})^{-1}(L)\right)$$

  where

  - $\mathbb{E}(X) := \mathrm{Tr}\left(U_{\theta_0}\rho_\infty \otimes |\psi\rangle\langle\psi|\, U_{\theta_0}^\dagger\, X\right)$ is the stationary state at θ_0
  - $L := \langle\psi|K|\psi\rangle$ is the conditional expectation of K onto the system

Fisher information = variance of generator

- The asymptotic Fisher information is $H(\theta_0) = 4V(K,K)$ with 'variance'

$$V(K,K) \quad := \quad \mathbb{E}\left(K^2\right) + 2\mathbb{E}\left(K \circ (\mathrm{Id} - T_{\theta_0})^{-1}(L)\right)$$

  where

  - $\mathbb{E}(X) := \mathrm{Tr}\left(U_{\theta_0}\rho_\infty \otimes |\psi\rangle\langle\psi| \, U_{\theta_0}^\dagger X\right)$ is the stationary state at θ_0
  - $L := \langle\psi|K|\psi\rangle$ is the conditional expectation of K onto the system

- Interpretation:
  - limit model is family of coherent states $|\sqrt{H/2}u\rangle = \exp(iu\,\mathbb{G}(K))|0\rangle$
  - for optimal estimation of u measure conjugate variable of $\mathbb{G}(K)$

Idea of the proof

1. Reduce to a semigroup property

$$\langle \psi_{u,n} \mid \psi_{v,n} \rangle = \left\langle \xi \,\middle|\, T^n_{u/\sqrt{u},v/\sqrt{n}}(\mathbf{1}) \,\middle|\, \xi \right\rangle,$$

where $T_{u,v}$ is a continuos family of contractions on $M(\mathbb{C}^d)$ such that $T_{0,0} = T$.

2. Expand
$T_{u/\sqrt{u},v/\sqrt{n}} = T + \frac{1}{\sqrt{n}} T_1 + \frac{1}{n} T_2 + o(n^{-1})$ and decompose $M(\mathbb{C}^d) = \mathbb{C}\mathbf{1} \oplus \mathcal{L}$ s. t.

   - $T(\mathcal{L}) \subset \mathcal{L}$ and T is a strict contraction on $\mathcal{L}$
   - $T_1(\mathcal{L}) \subset \mathcal{L}$ and $T_1(\mathbf{1}) \in \mathcal{L}$

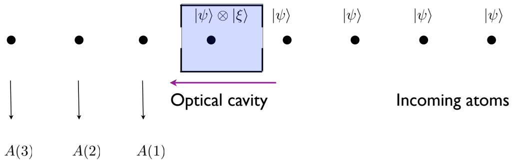
3. Use continuity and the spectral gap to expand the heighest eigenvector/eigenavalue of $T_{u/\sqrt{n},v/\sqrt{n}}$

$$T^n_{u/\sqrt{n},v/\sqrt{n}}(\mathbf{1}) \approx (1 + \lambda_2/n)^n \mathbf{1} \to \exp(\lambda_2)\mathbf{1}$$

where

$$\lambda_2 = \left[T_2 + T_1 \circ (\mathrm{Id} - T)^{-1} \circ T_1 \right]_{1,1}$$

Asymptotic normality for simple measurements



- Output state $|\psi_{u,n}\rangle := \left(S \circ U_{\theta_0+u/\sqrt{n}}\right)^n \left|\xi \otimes \psi^{\otimes n}\right\rangle$
- Measure the same observable A with $\mathbb{E}_{\theta_0}(A) = 0$ on each atom

Theorem

1. The Central Limit Theorem holds:

$$\bar{A}_n := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} A(k) \xrightarrow{\mathcal{D}} N(u\mu(A), V(A,A))$$

2. Estimator $\hat{u}_n := \bar{A}_n / \mu(A)$ with variance (inverse Fisher information)

$$\mathbb{E}\left[(\hat{u}_n - u)^2\right] \to \frac{V(A,A)}{\mu(A)^2}$$

Asymptotic normality for simple measurements

Variance and 'speed' of $\bar{A}_n$

$$
\begin{aligned}
V(A, A) &:= \mathbb{E}\left(A^2\right) + 2\mathbb{E}\left(A \otimes (\mathrm{Id} - T_{\theta_0})^{-1}(B)\right) \\
\mu(A) &:= \mathbb{E}\left(i[K, A \otimes \mathbf{1} + \mathbf{1} \otimes (\mathrm{Id} - T_{\theta_0})^{-1}(B)]\right)
\end{aligned}
$$

where $B := \langle\psi| U_{\theta_0}^{\dagger} A U_{\theta_0} |\psi\rangle$

Example: X-Y (spin-spin) interaction

Unitary interactions

$$U = \exp(i\theta(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y))$$

Input state

$$|\psi\rangle = a|0\rangle + b|1\rangle$$

Quantum Fisher information

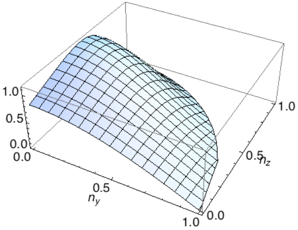
$$H = \frac{16|ab|^4}{(1-\cos\theta_0)(1-\cos\theta_0+4|ab|^2\cos\theta_0)}$$

Spin Measurement

in direction $\vec{n} = (n_x, n_y, n_z)$

Classical Fisher information

$$I(X) = \mu(X)^2/\sigma^2(X)$$



Singular point $\theta_0 = 0$: quantum Fisher information scales as n^2 !

Outlook

- Quantum Engineering needs Statistics!
- A variety of quantum statistical models are asymptotically normal
- Work in progress:
  - extension to continuous time and multiple parameters
  - general quantum Central Limit Theorem / Large Deviations
  - link to systems theory (engineering) and adaptive control

More information:

Quantum Statistics course (10 h)

http://maths.dept.shef.ac.uk/magic/course.php?id=181

Valparaiso Winter School on Stochastic Processes (6 h)

http://www.maths.nottingham.ac.uk/personal/pmzmig/preprints/Valparaiso.pdf

Lunteren Stochastics Meeting lectures (2h)

http://www.maths.nottingham.ac.uk/personal/pmzmig/Lunteren.pdf

References

L. Artiles, M. Guta and R. D. Gill
An invitation to quantum tomography
J. Royal Statist. Soc. B: Methodology 67, 109-134 (2005)

C. Butucea, M. Guta and L. Artiles
Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data
Ann. Statist. 35, 465-494, (2007)

M. Guta, J. Kahn
Local asymptotic normality for qubit states
Phys. Rev. A 73, 05218, (2006)

M. Guta, A. Jencova
Local asymptotic normality in quantum statistics
Commun. Math. Phys. 276, 341-379, (2007)

M. Guta, B. Janssens and J. Kahn
Optimal estimation of qubit states with continuous time measurements
Commun. Math. Phys. 277, 127-160, (2008)

J. Kahn, M. Guta
Local asymptotic normality for finite dimensional quantum systems
Commun. Math. Phys, 289, 597-652 (2009)

M. Guta, W. Kotlowski
Quantum learning: asymptotically optimal classification of qubit states
New J. Phys. 12 123032 (2010)

M. Guta
Fisher information and asymptotic normality in system identification for quantum Markov chains
Phys. Rev. A, 83, 062324 (2011)