# Quantum statistics:

# optimal state estimation

**Mădălin Guță**

School of Mathematical Sciences
University of Nottingham

*Quantum trajectories, parameter and state estimation*

*Toulouse 2017*

# Outline

- PART I: Basic notions of classical statistics
- PART II: Quantum statistics intro
- PART III: Quantum local asymptotic normality

# 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

## Basic notions

- 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 which was the underlying distribution
- Sufficient statistic: can data $X \sim \mathbb{P}_\theta$ be 'summarised' into a 'simpler' statistics $f(X)$ without losing information about the unknown parameter $\theta$ ?
- 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 $\Theta$ be a parameter space. A statistical model over $\Theta$ is a family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ of probability distributions on a measure space $(\mathcal{X}, \Sigma)$.

# Statistical models

## Definition

Let $\Theta$ be a parameter space. A statistical model over $\Theta$ 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:

$$\mathbb{P}_\theta^n([X_1 = x_1, \ldots, X_n = x_n]) = \prod_{i=1}^{n} \mathbb{P}_\theta([X_i = x_i]) = \theta^{\#1s} \cdot (1 - \theta)^{\#0s}$$

- **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 $\phi$ in plane. For each $\phi$ we get data $X \sim \mathcal{R}[p](x|\phi)$ where $\mathcal{R}[p]$ is the Radon transform

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

# Parametric estimation

## Problem

*Given*

- *a subset* $\Theta$ *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* $\Theta$ *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' $\Theta$, 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

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 (MSE) of estimator $\hat{\theta}(X) =$ variance $+$ bias

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

- estimator $\hat{\theta}(X)$ is called unbiased if $\mathbb{E}_\theta(\hat{\theta}(X)) = \theta$ for all $\theta$.
  If $\hat{\theta}$ is unbiased then the mean square error is equal to $V(\hat{\theta})$.

# Unbiased estimators

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 (MSE) of estimator $\hat{\theta}(X) =$ variance $+$ bias

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

- estimator $\hat{\theta}(X)$ is called unbiased if $\mathbb{E}_\theta(\hat{\theta}(X)) = \theta$ for all $\theta$.
  If $\hat{\theta}$ is unbiased then the mean square error is equal to $V(\hat{\theta})$.

## Example (Gaussian shift)

Let $Y_1, \ldots, Y_n$ be i.i.d. normally distributed with $\mathbb{P}_\theta = N(\theta, V)$.

Then $\hat{\theta}_n := \bar{Y}_n \equiv (\sum_i Y_i)/n$ is an unbiased estimator of $\theta$ with MSE

$$\mathbb{E}_\theta\left[(\hat{\theta}_n-\theta)^2\right] = \mathbb{E}_\theta\left[\left(\frac{1}{n}\sum_{i=1}^n (Y_i-\theta)\right)^2\right] = \frac{1}{n}V$$

# Fisher information and the Cramér-Rao bound

- 'smooth' parametric statistical model

  $\mathbb{P}_\theta$ has probability density $p_\theta = \frac{d\mathbb{P}_\theta}{d\mu}$ which is smooth with respect to $\theta \in \Theta \subset \mathbb{R}^k$

- score functions $\dot{\ell}_{\theta,i} := \partial \log p_\theta / \partial \theta_i, \qquad i = 1, \ldots, k$

- Fisher information matrix

$$I_{i,j}(\theta) := \mathbb{E}_\theta(\dot{\ell}_{\theta,i}\dot{\ell}_{\theta,j}) = \int_{x:p_\theta(x)>0} p_\theta^{-1}(x) \frac{\partial p_\theta}{\partial \theta_i}(x) \frac{\partial p_\theta}{\partial \theta_j}(x) \mu(dx)$$

# Fisher information and the Cramér-Rao bound

- 'smooth' parametric statistical model

  $\mathbb{P}_\theta$ has probability density $p_\theta = \frac{d\mathbb{P}_\theta}{d\mu}$ which is smooth with respect to $\theta \in \Theta \subset \mathbb{R}^k$

- score functions $\dot{\ell}_{\theta,i} := \partial \log p_\theta / \partial \theta_i, \qquad i = 1, \ldots, k$

- Fisher information matrix

$$I_{i,j}(\theta) := \mathbb{E}_\theta(\dot{\ell}_{\theta,i}\dot{\ell}_{\theta,j}) = \int_{x:p_\theta(x)>0} p_\theta^{-1}(x)\frac{\partial p_\theta}{\partial \theta_i}(x)\frac{\partial p_\theta}{\partial \theta_j}(x)\mu(dx)$$

**Theorem (Cramér-Rao)**

*Let $\hat{\theta}$ be an unbiased estimator of $\theta$. 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 for one-dimensional $\theta$: by Cauchy-Schwarz in $L^2(\Omega, \Sigma, \mu)$

$$\mathrm{Var}(\hat{\theta}) \cdot I(\theta) = \mathbb{E}_\theta((\hat{\theta}-\theta)^2) \cdot \mathbb{E}_\theta(\dot{\ell}_\theta^2) \geq \left|\mathbb{E}_\theta((\hat{\theta}-\theta)\dot{\ell}_\theta)\right|^2 = 1$$

# 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:
  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 $p_\theta$ and $p_{\theta+d\theta}$ is determined by the Fisher information

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

- randomisation: $T : L^1(\Omega, \sigma, \mu) \to L^1(\Omega', \sigma', \mu')$, linear, maps pdfs to pdfs (e.g. randomised statistic, Markov kernel). Let $p'_\theta := T(p_\theta)$. Then

$$h(p'_{\theta_1}, p'_{\theta_2}) \leq h(p_{\theta_1}, p_{\theta_2}) \qquad \text{and} \qquad I'(\theta) \leq I(\theta)$$

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

# Example: repeated coin toss

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

- $\hat{\theta}_n := \bar{X}_n \equiv (\sum_{i=1}^n X_i)/n$ is an unbiased estimator of $\theta$. 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 Fisher information is

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta[\dot{\ell}_\theta^2] = \theta \cdot \left(\frac{d \log \theta}{d\theta}\right)^2 + (1-\theta) \cdot \left(\frac{d \log(1-\theta)}{d\theta}\right)^2 \\ &= \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)} = \frac{1}{Var(X)} \end{aligned}$$

- $\hat{\theta}_n$ attains the Cramér-Rao bound:

$$\mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right] = \frac{1}{n} Var(X) = \frac{\theta(1-\theta)}{n} = \frac{1}{nI(\theta)}$$

# 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 $X_1, \ldots, X_n$ 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:

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

## Theorem

*Under regularity conditions, the maximum likelihood estimator*

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

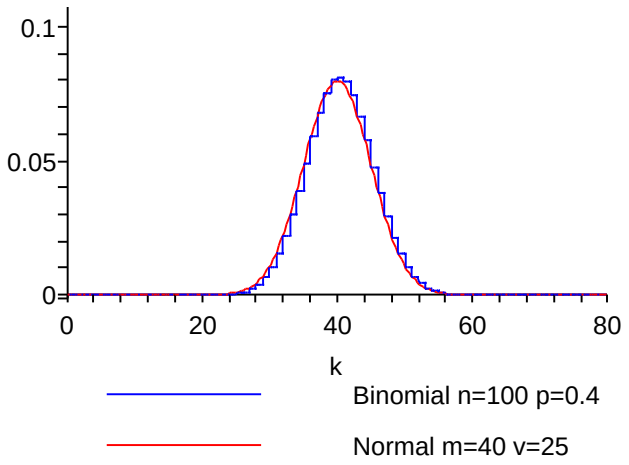
*is asymptotically efficient.*

# Example: repeated coin toss

Let $X_1, \ldots, X_n$ be i.i.d. with $X_i \sim \mathrm{Bernoulli}(\theta)$

- Estimator $\hat{\theta}_n := \bar{X}_n \equiv (\sum_{i=1}^n X_i)/n$ has binomial distribution Bin$(n, \theta)$

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




- By the Central Limit Theorem we have

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

- Thus $\bar{X}_n$ is asymptotically efficient (and is equal to the max. lik. estimator).

# Local asymptotic normality for coin toss

- CLT: $\hat{\theta}_n$ is approximately distributed as

$$\hat{\theta}_n \approx \theta + \frac{1}{\sqrt{n}}\epsilon$$

  with 'error' $\epsilon \sim N(0, \theta(1-\theta))$ whose variance depends on $\theta$.

- A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters $\theta$ ....

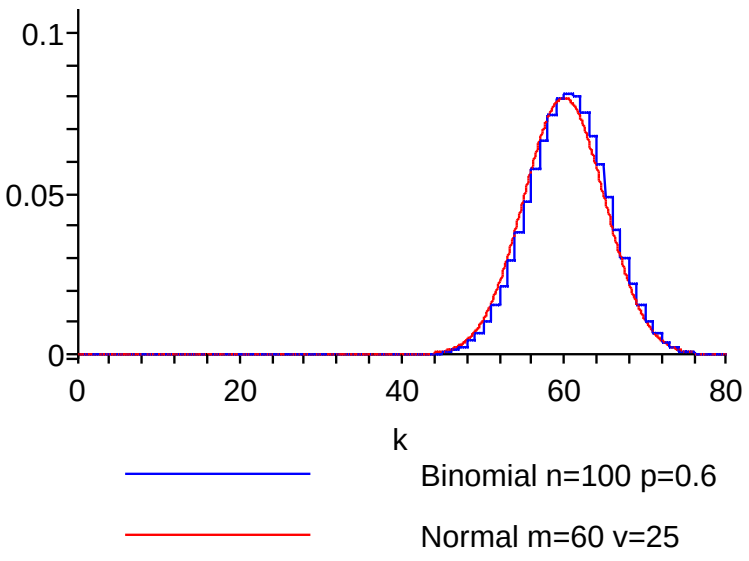
# Local asymptotic normality for coin toss

- CLT: $\hat{\theta}_n$ is approximately distributed as

$$\hat{\theta}_n \approx \theta + \frac{1}{\sqrt{n}}\epsilon$$

with 'error' $\epsilon \sim N(0, \theta(1-\theta))$ whose variance depends on $\theta$.

- A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters $\theta$ ....



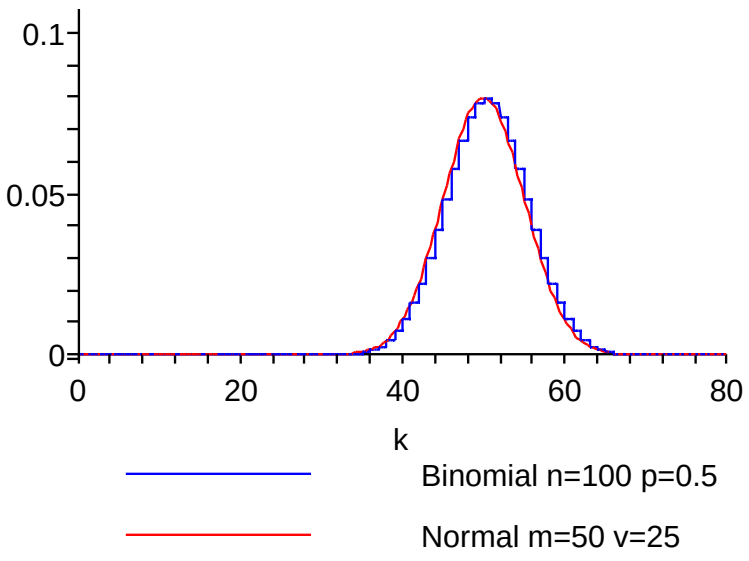
# Local asymptotic normality for coin toss

- CLT: $\hat{\theta}_n$ is approximately distributed as

$$\hat{\theta}_n \approx \theta + \frac{1}{\sqrt{n}}\epsilon$$

with 'error' $\epsilon \sim N(0, \theta(1-\theta))$ whose variance depends on $\theta$.

- A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters $\theta$ ....



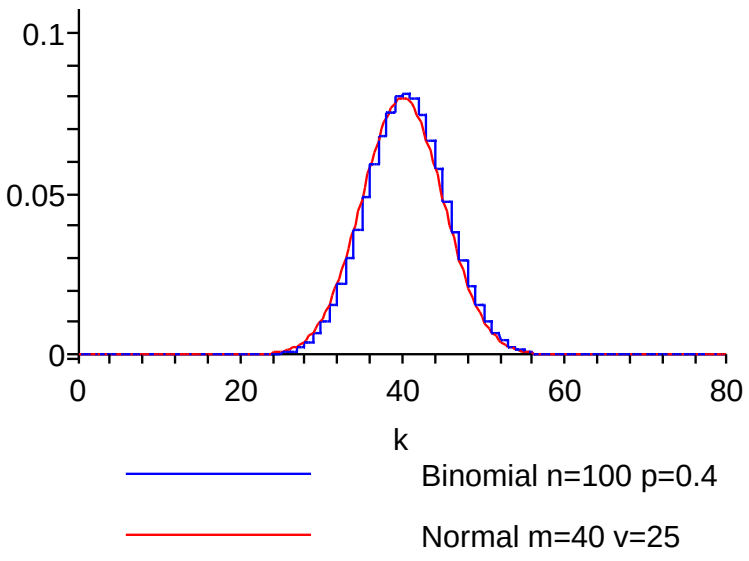
# Local asymptotic normality for coin toss

- CLT: $\hat{\theta}_n$ is approximately distributed as

$$\hat{\theta}_n \approx \theta + \frac{1}{\sqrt{n}}\epsilon$$

with 'error' $\epsilon \sim N(0, \theta(1-\theta))$ whose variance depends on $\theta$.

- A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters $\theta$ ....



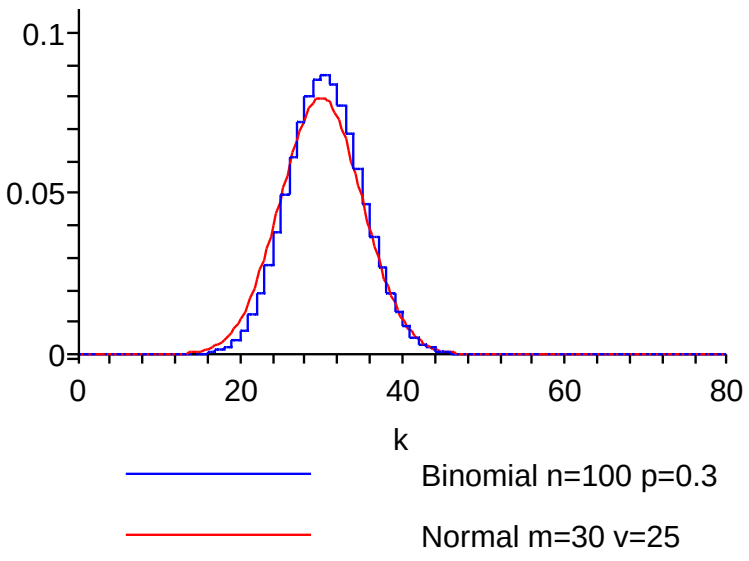
# Local asymptotic normality for coin toss

- CLT: $\hat{\theta}_n$ is approximately distributed as

$$\hat{\theta}_n \approx \theta + \frac{1}{\sqrt{n}}\epsilon$$

with 'error' $\epsilon \sim N(0, \theta(1-\theta))$ whose variance depends on $\theta$.

- A closer look shows that a Gaussian with a fixed variance is a good fit for the binomial for a whole interval of parameters $\theta$ ....



# Local asymptotic normality for coin toss

- For large $n$, $\theta$ can be 'localised' in a region of size $n^{-1/2+\epsilon}$ centred at some $\theta_0$
- Switch to local model: $\theta = \theta_0 + u/\sqrt{n}$ with unknown local parameter $u$.
- Rewrite $\hat{\theta}_n = \theta_0 + \hat{u}_n/\sqrt{n}$ in terms of sufficient statistic

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

**Lemma (LAN with weak convergence)**

"**The binomial model with local parameter $u$ converges to a Gaussian shift model**".

*For any local parameter $u$ the convergence in distribution holds*

$$\hat{u}_n \xrightarrow{\mathcal{L}} N\left(u, I_{\theta_0}^{-1}\right), \qquad I_{\theta_0} = \frac{1}{\theta_0(1-\theta_0)}$$

# Local asymptotic normality for coin toss

- For large $n$, $\theta$ can be 'localised' in a region of size $n^{-1/2+\epsilon}$ centred at some $\theta_0$
- Switch to local model: $\theta = \theta_0 + u/\sqrt{n}$ with unknown local parameter $u$.
- Rewrite $\hat{\theta}_n = \theta_0 + \hat{u}_n/\sqrt{n}$ in terms of sufficient statistic

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

**Lemma (LAN with weak convergence)**

"**The binomial model with local parameter $u$ converges to a Gaussian shift model**".

*For any local parameter $u$ the convergence in distribution holds*

$$\hat{u}_n \xrightarrow{\mathcal{L}} N\left(u, I_{\theta_0}^{-1}\right), \qquad I_{\theta_0} = \frac{1}{\theta_0(1-\theta_0)}$$

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

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

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

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

# 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

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

# 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

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

- Weak convergence:

  convergence in distribution of likelihood ratio process (sufficient statistic)

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

# 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

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

- Strong convergence (Le Cam):

  there exist randomizations $T_n, S_n$ such that

$$\lim_{n \to \infty} \sup_{\|u\| < a} \left\| T_n \, p^n_{\theta_0 + u/\sqrt{n}} - n(u, I^{-1}_{\theta_0}) \right\|_1 = 0$$

  and

$$\lim_{n \to \infty} \sup_{\|u\| < a} \left\| p^n_{\theta_0 + u/\sqrt{n}} - S_n \, n(u, I^{-1}_{\theta_0}) \right\|_1 = 0$$

  where $n(u, I^{-1}_{\theta_0})$ is the pdf of $N(u, I^{-1}_{\theta_0})$.

# PART II: Quantum statistics intro

# Quantum mechanics up to the 60's

- Q.M. predicts probability distributions of measurement outcomes
- Perform measurements on huge ensembles
- Observed frequencies = probabilities

**Old Paradigm**

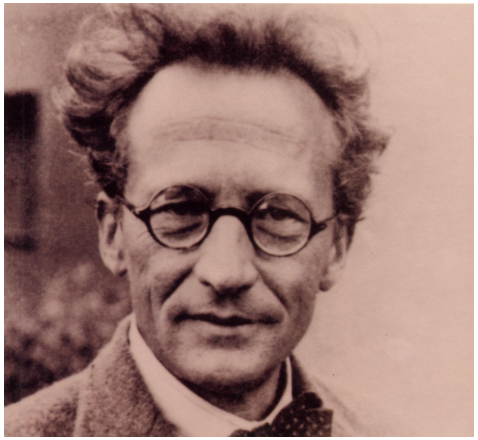
It makes no sense to talk about individual quantum systems

E. Schrödinger

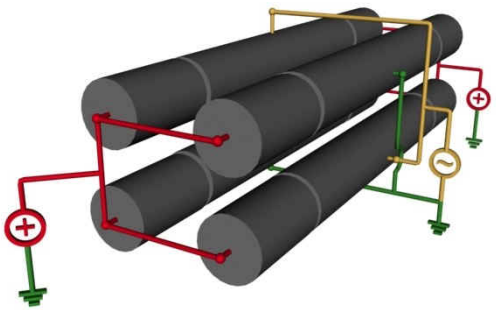
["Are there quantum jumps ?", British J.Phil. Science 1952]

"We are not experimenting with single particles, any more than we can raise Ichtyosauria in the zoo.

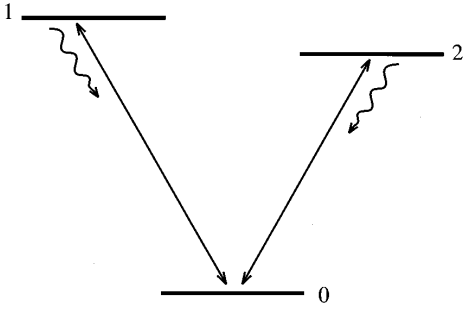
We are scrutinizing records of events long after they have happened."



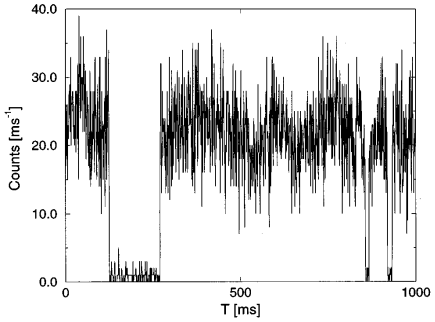
# Are there quantum jumps ?



Paul ion trap




3 level atom driven by 2 lasers




Recorded fluorescence signal from 1 ion

- First experiments with individual quantum systems
- Measurements with stochastic outcomes
- Stochastic Schrödinger equations

# Influx of mathematical ideas in the 70's

E. B. Davies

V. P. Belavkin

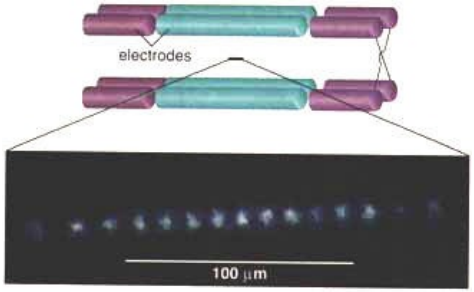
A. S. Holevo

C. W. Helstrom

- Probability
  what is the nature of quantum noise
- Filtering and Control Theory
  how to monitor and control quantum systems in noisy environments in real time
- Information Theory
  how to encode, transmit and decode quantum information
- Statistics and metrology
  how to learn unknown parameters from measurement outcomes

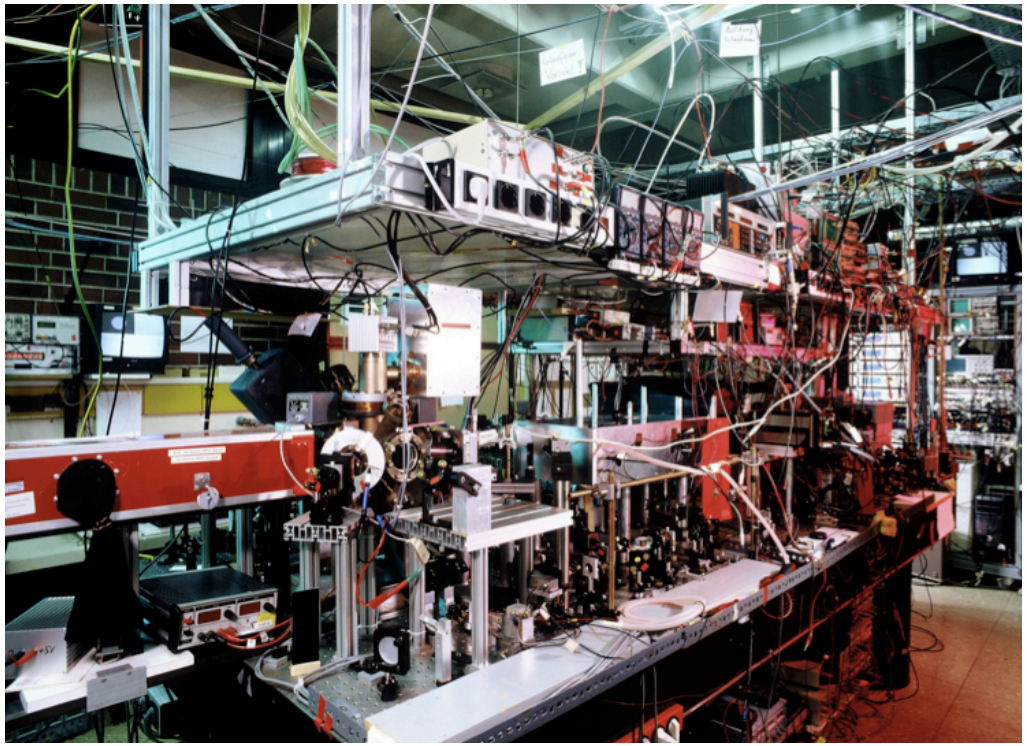
# Quantum statistics goes viral

- Problem: Quantum state estimation
  - Quantum computation goal: create specific states of multiple "qubits" (e.g. ions)
  - Validation: statistical estimation from measurement outcomes




[Häffner *et al*, Nature 2005]

- $4^8 - 1 = 65\,535$ parameters
- $3^8 \times 100 = 656\,100$ measurements
- 10 hours measurement time
- days of computer time



Rainer Blatt's Lab, Innsbruck

"quantum computer" with 8 qubits (ions)

# Quantum states

- Complex Hilbert space of 'wave functions' $\mathcal{H} = \mathbb{C}^d, L^2(\mathbb{R})$...
- State = preparation: complex density matrix $\rho$ on $\mathcal{H}$
  - $\rho = \rho^*$ (selfadjoint)
  - $\rho \geq 0$ (positive)
  - $\mathrm{Tr}(\rho) = 1$ (normalised)

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1} & \rho_{d2} & \cdots & \rho_{dd} \end{pmatrix}$$

- Convex space of states
  - extremals (pure states) : one dimensional projection $P_\psi = |\psi\rangle\langle\psi| = \psi\psi^*$ with $\|\psi\| = 1$
  - Mixed state: convex combination of pure states

$$\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$$

- Natural distances:
  - operator norm: $\|\rho_1 - \rho_2\| := \lambda_{\max}(|\rho_1 - \rho_2|)$
  - trace norm: $\|\rho_1 - \rho_2\|_1 := \mathrm{Tr}(|\rho_1 - \rho_2|)$
  - norm-two: $\|\rho_1 - \rho_2\|_2^2 := \mathrm{Tr}(|\rho_1 - \rho_2|^2)$
  - Bures distance: $d_b^2(\rho_1, \rho_2) := 2(1 - \mathrm{Tr}\sqrt{\rho_1^{1/2}\,\rho_2\rho_1^{1/2}})$

# Simple measurements and observables

- Measurement setting given by an orthonormal basis $\mathbf{s} := \{|e_1\rangle, \ldots, |e_d\rangle\}$ in $\mathbb{C}^d$
- Outcome of measurement is a random index of a basis element $O \in \{1, \ldots, d\}$
- Probability distribution: if system is prepared state $\rho$

$$\mathbb{P}[O = i] = \langle e_i | \rho | e_i \rangle = \rho_{ii}$$

# Simple measurements and observables

- Measurement setting given by an orthonormal basis $\mathbf{s} := \{|e_1\rangle, \ldots, |e_d\rangle\}$ in $\mathbb{C}^d$
- Outcome of measurement is a random index of a basis element $O \in \{1, \ldots, d\}$
- Probability distribution: if system is prepared state $\rho$

$$\mathbb{P}[O = i] = \langle e_i|\rho|e_i\rangle = \rho_{ii}$$

- Observable: selfadjoint operator $A$ on $\mathcal{H}$

$$A = \int_{\sigma(A)\subset\mathbb{R}} a\, P(da) \qquad (A = \sum_j a_j P_j)$$

- Probabilistic interpretation: measuring $A$ gives random outcome $\mathbf{A} \in \{a_j\}$

$$\mathbb{P}[\mathbf{A} = a_j] = p_j = \mathrm{Tr}(\rho P_j)$$

# General measurements as “quantum-to-classical randomisations"

## Definition

A measurement on $\mathcal{H}$ with outcomes in $(\Omega, \Sigma)$ is a linear map

$$M : \mathcal{T}_1(\mathcal{H}) \to L^1(\Omega, \Sigma, \mathbb{P})$$

such that $p_\rho := M(\rho)$ is a probability density for each state $\rho$.

Any $M$ is of the form

$$\mathbb{P}_\rho(E) = \mathrm{Tr}(\rho\, m(E))$$

for some Positive Operator valued Measure (POVM) $\{m(E) : E \in \Sigma\}$.

- Naimark’s Theorem

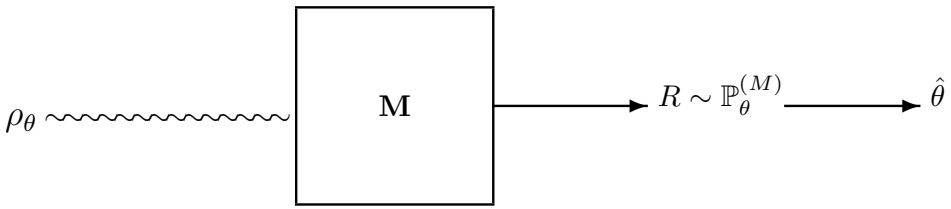
  Any measurement can be realised indirectly by ‘usual’ projection measurement in an ‘enlarged’ Hilbert space

# Set-up of quantum estimation problems

- Quantum statistical model over $\Theta$:

$$\mathcal{Q} = \{\rho_\theta \ : \ \theta \in \Theta\}$$

- Estimation procedure: measure state $\rho_\theta$ and devise estimator $\hat{\theta} = \hat{\theta}(R)$

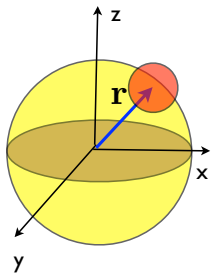



- Measurement design:
  - which classical model $\mathcal{P}^{(M)} = \{\mathbb{P}_\theta^{(M)} : \theta \in \Theta\}$ is 'best' ?
  - trade-off between incompatible observables
  - optimal measurement depends on statistical problem

# Example: spin / two-level ion / qubit tomography

- Any state on $\mathbb{C}^2$ is parametrized by a 3-D Bloch vector $\mathbf{r} = (r_x, r_y, r_z)$ with $\|\mathbf{r}\| \leq 1$

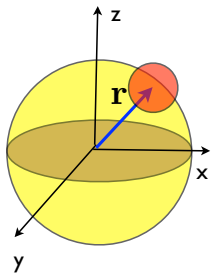
$$\rho_{\mathbf{r}} = \frac{1}{2} \begin{pmatrix} 1 + r_z & r_x - ir_y \\ r_x + ir_y & 1 - r_z \end{pmatrix}$$



# Example: spin / two-level ion / qubit tomography

- Any state on $\mathbb{C}^2$ is parametrized by a 3-D Bloch vector $\mathbf{r} = (r_x, r_y, r_z)$ with $\|\mathbf{r}\| \leq 1$

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




- 3 standard measurement bases corresponding to $s = x, y, z$ spin observables

$$\underbrace{|e_x^{\pm}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}}_{s=x} \qquad \underbrace{|e_y^{\pm}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm i \end{pmatrix}}_{s=y} \qquad \underbrace{|e_z^{+}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |e_z^{-}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{s=z}$$

- Probability distributions: $\mathbb{P}(o = \pm|s) = \frac{1 \pm r_s}{2}, \qquad s = x, y, z$

- $n$ measurement repetitions $\longrightarrow$ counts $\{N(\pm|x), N(\pm|y), N(\pm|z)\} \longrightarrow$ (LS) estimator

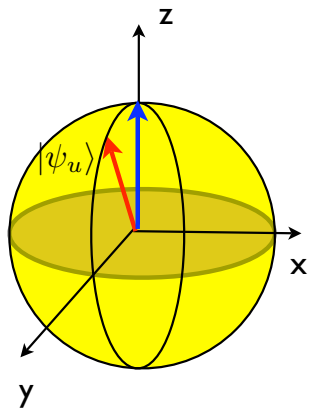
$$\widehat{\rho}_n := \rho_{\hat{\mathbf{r}}}, \qquad \widehat{r}_{x,y,z} := \frac{N(+|x,y,z) - N(-|x,y,z)}{n}$$

Boundary/positivity problem: for pure (rank one) states, estimator may not be physical (positive)

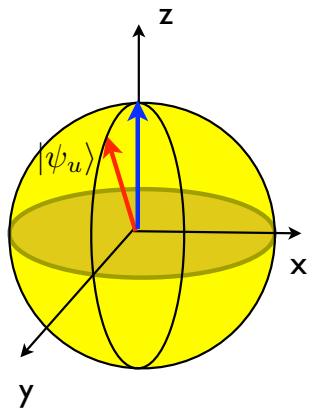
# Example: estimating the direction of the spin vector

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

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



# Example: estimating the direction of the spin vector



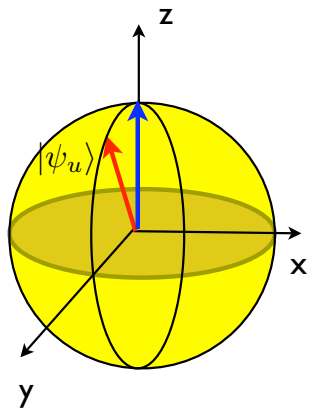

- One-dim. model: (small) rotation of $|\uparrow\rangle$

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

- 'Most informative' spin observable is $\sigma_y$

$$\mathbb{E}(\sigma_y) = \sin(2u) \approx 2u$$

# Example: estimating the direction of the spin vector




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

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

- 'Most informative' spin observable is $\sigma_y$

$$\mathbb{E}(\sigma_y) = \sin(2u) \approx 2u$$

- Two parameter model $|\psi_{u_x,u_y}\rangle = \exp(i(u_y\sigma_x - u_x\sigma_y))|\uparrow\rangle$

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

# Quantum Cramér-Rao bound*

**Theorem** [Helstrom, Holevo, Belavkin]

Let $\mathcal{Q} = \{\rho^\theta : \theta \in \mathbb{R}^k\}$ be a 'smooth' quantum model.

For any unbiased measurement $M$ with outcome $\hat{\theta} \in \mathbb{R}^k$

$$\mathrm{Var}(\hat{\theta}) \geq I^{(M)}(\theta)^{-1} \geq F(\theta)^{-1}$$

where $I^{(M)}(\theta)^{-1}$ is the Fisher information of $\hat{\theta}$.

- Quantum Fisher information matrix

$$F(\theta)_{i,j} := \mathrm{Tr}(\rho_\theta \mathcal{L}_{\theta,i} \circ \mathcal{L}_{\theta,j})$$

- Symmetric logarithmic derivatives: $\frac{\partial \rho_\theta}{\partial \theta_j} = \rho_\theta \circ \mathcal{L}_{\theta,j}$

*= several C.R. bounds exist depending on symmetrisation

# Proof (projection valued measurements)

Measure observable $X$ and get result $\mathbf{X} \equiv \hat{\theta} \sim \mathbb{P}_\theta$

- Hilbert spaces $L^2(\rho_\theta)$ and $L^2(\mathbb{R}, \mathbb{P}_\theta)$

$$\langle A, B \rangle_\theta := \mathrm{Tr}(\rho_\theta A \circ B) \qquad \langle f(\mathbf{X}), g(\mathbf{X}) \rangle_\theta = \mathbb{E}_\theta(f(\mathbf{X})g(\mathbf{X}))$$

- Fisher informations

$$\begin{aligned} F(\theta) &:= \mathrm{Tr}(\rho_\theta \mathcal{L}_\theta^2) = \|\mathcal{L}_\theta\|_\theta^2 \\ I^{(M)}(\theta) &:= \mathbb{E}_\theta(\ell_\theta^2(\mathbf{X})) = \|\ell_\theta(\mathbf{X})\|_\theta^2 \end{aligned}$$

- Isometry

$$\begin{aligned} I : L^2(\mathbb{R}, \mathbb{P}_\theta) &\to L^2(\rho_\theta) \\ f(\mathbf{X}) &\mapsto f(X) \end{aligned}$$

$$\mathbb{E}_\theta(f(\mathbf{X})g(\mathbf{X})) = \mathrm{Tr}(\rho_\theta f(X) \circ g(X)) = \langle f(X), g(X) \rangle_\theta$$

# Proof for projection valued measurements

- The projection of $\mathcal{L}_\theta$ onto $IL^2(\mathbb{R}, \mathbb{P}_\theta)$ is $\ell_\theta(X)$

  Indeed for every $f \in L^2(p_\theta)$

$$\langle f(\mathbf{X}), \ell_\theta(\mathbf{X})\rangle_\theta = \frac{d\mathbb{E}_\theta(f)}{d\theta} = \mathrm{Tr}\left(\frac{d\rho_\theta}{d\theta} f(X)\right) = \langle f(X), \mathcal{L}_\theta\rangle_\theta$$

  $\implies \|\ell_\theta\|_\theta^2 \leq \|\mathcal{L}_\theta\|_\theta^2$

$\square$

- Bound achieved (locally) at $\theta_0$ by $X = \theta_0 \mathbf{1} + \frac{\mathcal{L}_{\theta_0}}{H(\theta_0)}$

  - $\mathbb{E}_\theta X = \theta_0 + \frac{\mathrm{Tr}(\rho_\theta \mathcal{L}_{\theta_0})}{F(\theta_0)} = \theta_0 + \frac{\mathrm{Tr}(\rho_{\theta_0} \mathcal{L}_{\theta_0})}{F(\theta_0)} + \Delta\theta \frac{\mathrm{Tr}(\rho'_{\theta_0} \mathcal{L}_{\theta_0})}{F(\theta_0)} + O(\Delta\theta^2)$
    $= \theta_0 + \Delta\theta + O(\Delta\theta^2) = \theta + O(\Delta\theta^2)$

  - $Var_{\theta_0}(X) = \mathbb{E}_{\theta_0}\left[(X - \mathbb{E}_{\theta_0} X)^2\right] = \frac{\mathrm{Tr}(\rho_{\theta_0} \mathcal{L}_{\theta_0}^2)}{F^2(\theta_0)} = \frac{1}{F_{\theta_0}}$

# Trade-off between parameters

- One-dimensional model: C.R. bound can be achieved asymptotically
  1. measure fraction $\tilde{n} \ll n$ of systems to obtain rough estimator $\theta_0$
  2. measure $\mathcal{L}_{\theta_0}^{(n)} := \mathcal{L}_{\theta_0} \otimes \mathbf{1} \otimes \cdots \otimes \mathbf{1} + \cdots + \mathbf{1} \otimes \mathbf{1} \otimes \cdots \otimes \mathcal{L}_{\theta_0}$
  3. set $\hat{\theta}_n := \theta_0 + \mathbf{L}_{\theta_0}^{(n)} / nF(\theta_0)$

# Trade-off between parameters

- One-dimensional model: C.R. bound can be achieved asymptotically
  1. measure fraction $\tilde{n} \ll n$ of systems to obtain rough estimator $\theta_0$
  2. measure $\mathcal{L}_{\theta_0}^{(n)} := \mathcal{L}_{\theta_0} \otimes \mathbf{1} \otimes \cdots \otimes \mathbf{1} + \cdots + \mathbf{1} \otimes \mathbf{1} \otimes \cdots \otimes \mathcal{L}_{\theta_0}$
  3. set $\hat{\theta}_n := \theta_0 + \mathbf{L}_{\theta_0}^{(n)} / nF(\theta_0)$
- Multi-dimensioanal model: $F(\theta)$ is achievable iff

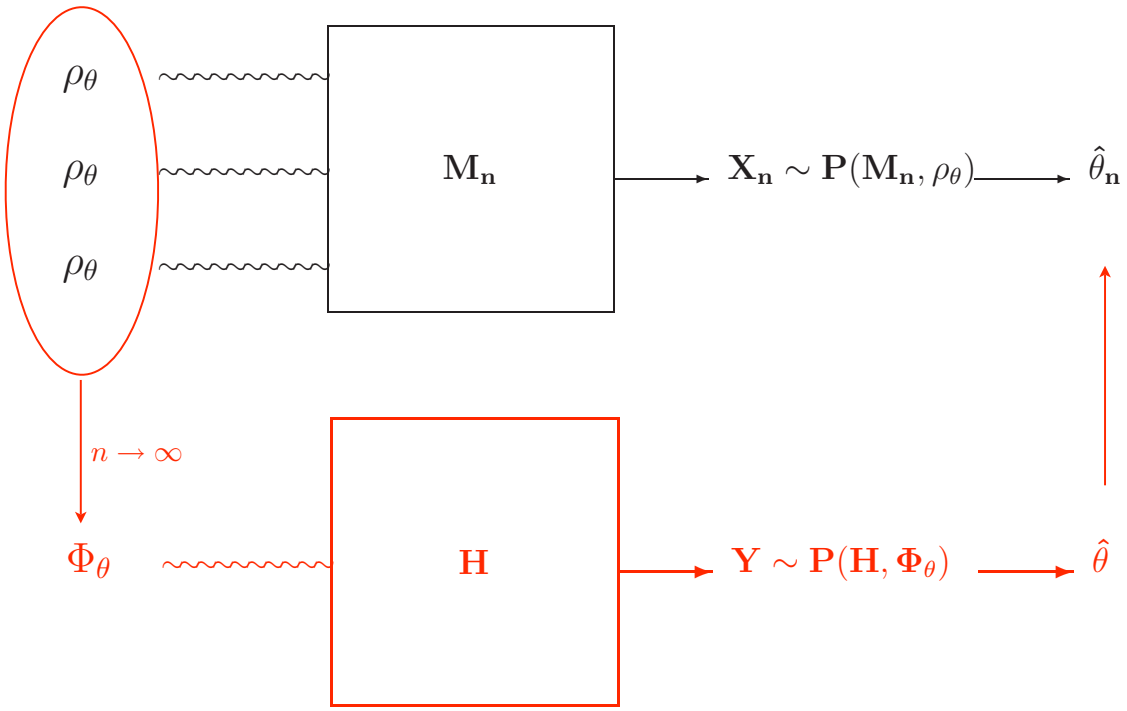
$$\mathrm{Tr}(\rho_\theta[\mathcal{L}_{\theta,j},\, \mathcal{L}_{\theta,i}]) = 0, \qquad \forall 1 \leq i, j \leq k$$

- Trade-off between estimation of different coordinates
- Optimal measurement depends on loss function

# PART III: Quantum local asymptotic normality

# Optimal estimation using local asymptotic normality

How to optimally estimate $\theta$ from an i.i.d. ensemble of states?




[L. Le Cam]

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

# Quantum Gaussian states

- Quantum 'particle' with canonical observables $Q, P$ on $\mathcal{H} = L^2(\mathbb{R})$

$$QP - PQ = i\mathbf{1} \qquad \text{(Heisenberg's commutation relations)}$$

# Quantum Gaussian states

- Quantum 'particle' with canonical observables $Q, P$ on $\mathcal{H} = L^2(\mathbb{R})$

$$QP - PQ = i\mathbf{1} \qquad \text{(Heisenberg's commutation relations)}$$

- Centred Gaussian state $\Phi$

$$\mathrm{Tr}\left(\Phi \exp(-ivQ - iuP)\right) = \exp\left(-\frac{1}{2}\begin{pmatrix} u & v \end{pmatrix} V \begin{pmatrix} u \\ v \end{pmatrix}\right)$$

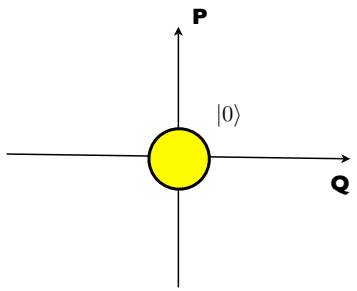
with 'covariance matrix' $V$ satisfying the uncertainty principle

$$\mathrm{Det}(V) = \begin{vmatrix} \mathrm{Tr}(\Phi Q^2) & \mathrm{Tr}(\Phi Q \circ P) \\ \mathrm{Tr}(\Phi Q \circ P) & \mathrm{Tr}(\Phi P^2) \end{vmatrix} \geq \frac{1}{4}$$

# Examples

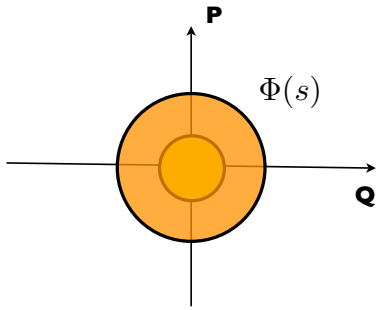
- Vaccum state $|0\rangle$

  $V = \mathrm{Diag}(\frac{1}{2},\ \frac{1}{2})$



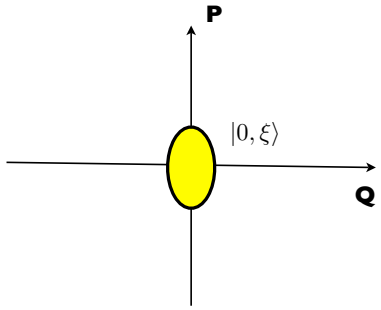

- Thermal equilibrium state $\Phi(s)$

  $V = \mathrm{Diag}(\frac{s}{2},\ \frac{s}{2})$




- Squeezed state $|0, \xi\rangle$

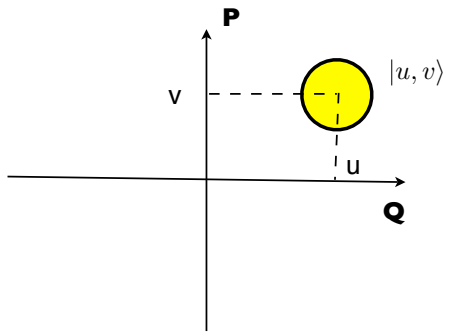
  $V = \mathrm{Diag}(\frac{e^{-\xi}}{2},\ \frac{e^{\xi}}{2})$



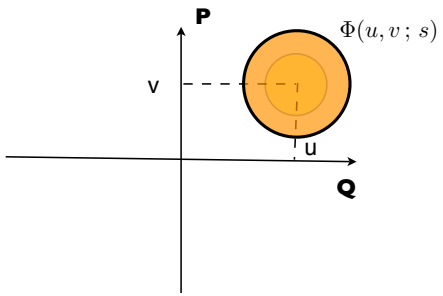
# Quantum Gaussian shift model(s)

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

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




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




$Q$ is the optimal measurement for estimating $u = \langle Q \rangle$

$P$ is the optimal measurement for estimating $v = \langle P \rangle$

How to optimally estimate both displacement parameters ?

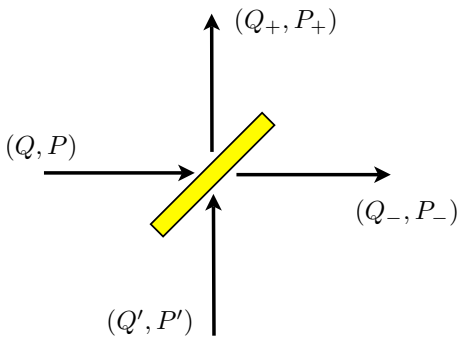
# Optimal measurement for Gaussian shift

- Oscillator $(Q, P)$ in state $|u, v\rangle$
- Oscillator $(Q', P')$ in vacuum state $|0\rangle$

# Optimal measurement for Gaussian shift

- Oscillator $(Q, P)$ in state $|u, v\rangle$
- Oscillator $(Q', P')$ in vacuum state $|0\rangle$
- Noisy coordinates commute: $[Q_+, P_-] = 0$

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

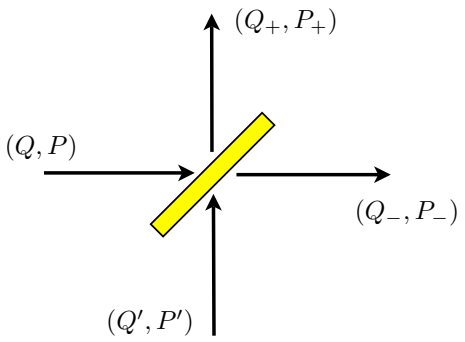



- Heterodyne measurement $(Q_+, P_-)$ gives estimator $(\hat{u}, \hat{v}) \sim N((u, v), \mathbf{1})$

# Optimal measurement for Gaussian shift

- Oscillator $(Q, P)$ in state $|u, v\rangle$
- Oscillator $(Q', P')$ in vacuum state $|0\rangle$
- Noisy coordinates commute: $[Q_+, P_-] = 0$

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




- Heterodyne measurement $(Q_+, P_-)$ gives estimator $(\hat{u}, \hat{v}) \sim N((u, v), \mathbf{1})$

**Theorem**

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

# Gaussian approximation (QCLT, CSS)

- n identically prepared spins

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

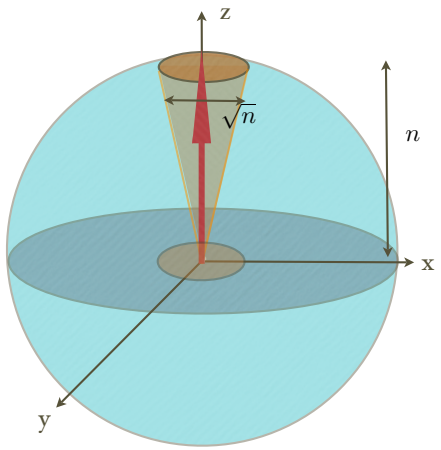
- Collective observables $L_{x,y,z} := \sum_{i=1}^{n} \sigma_{x,y,z}^{(i)}$

- Quantum Central Limit Theorem

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

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

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




[Holstein and Primakoff P.R. 1940] [Radcliffe J.Phys. A 1971]

# Gaussian approximation (QCLT, CSS)

- n identically prepared spins

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

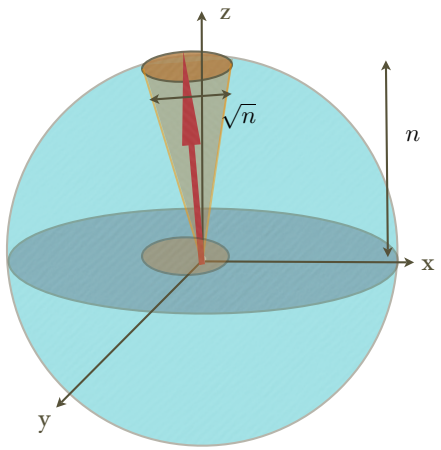
- Collective observables $L_{x,y,z} := \sum_{i=1}^{n} \sigma_{x,y,z}^{(i)}$

- Quantum Central Limit Theorem

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

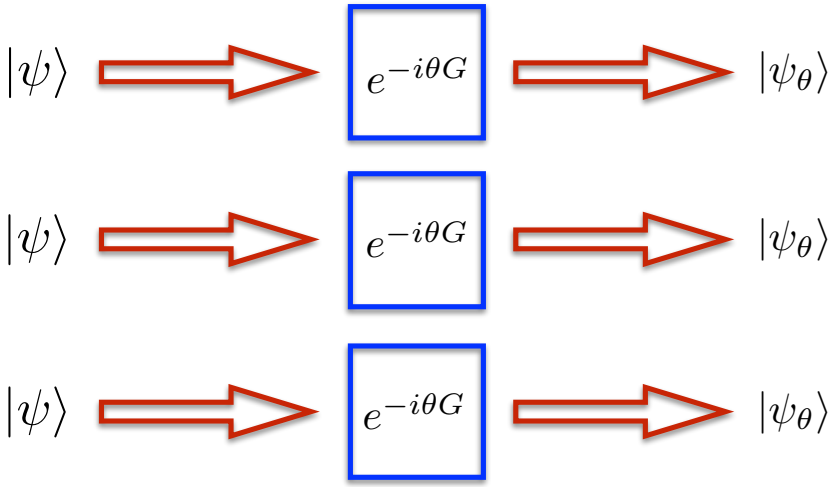
$$\frac{L_y}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(2v, 1)$$

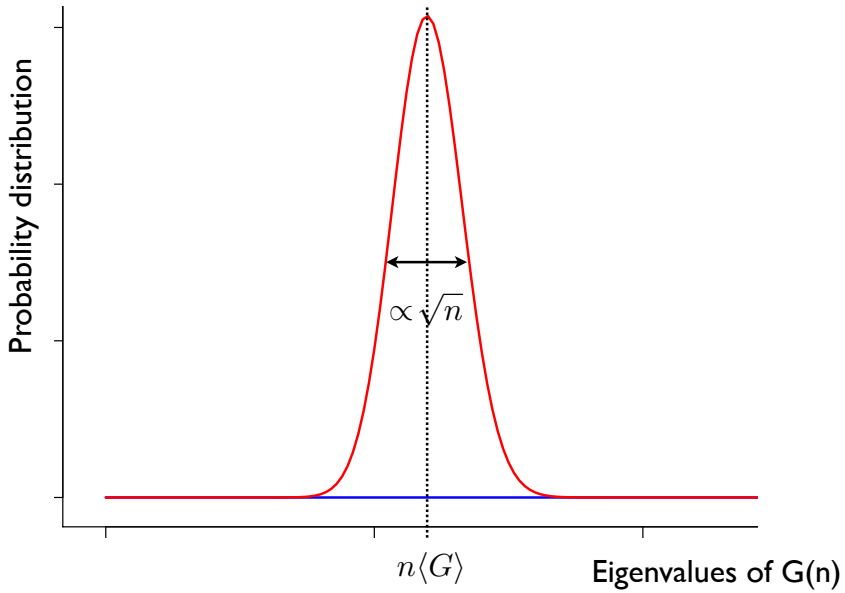
$$\left[ \frac{L_x}{\sqrt{n}}, \frac{L_y}{\sqrt{n}} \right] = \frac{2i}{n} L_z \xrightarrow{l.l.n.} 2i\mathbf{1}$$




[Holstein and Primakoff P.R. 1940] [Radcliffe J.Phys. A 1971]

# Standard scaling as Central Limit behaviour




- Quantum Fisher information

$$F(\theta) \quad = \quad 4\left\|\frac{d\psi_\theta}{d\theta}\right\|^2 = 4\mathrm{Var}_\psi(G) = 4\left\langle\psi\,\middle|\,G^2\,\middle|\,\psi\right\rangle$$

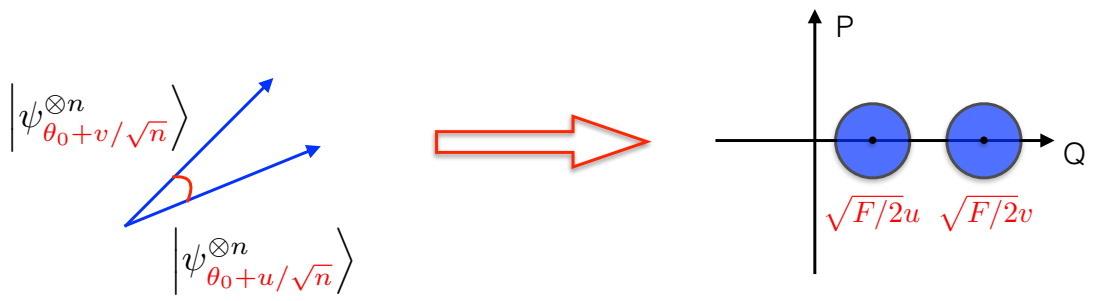
- total generator $G(n) := G^{(1)} + \cdots + G^{(n)}$
- Standard scaling: quantum Fisher information scales linearly in $n$

$$F_n(\theta) = 4\mathrm{Var}(G(n)) = 4n\mathrm{Var}(G)$$

# Convergence to Gaussian model for i.i.d. ensembles

- Quantum data: ensemble of $n$ identically prepared systems

$$|\psi_\theta\rangle^{\otimes n} := \left(e^{i\theta G}|\psi\rangle\right)^{\otimes n}, \qquad \langle\psi|G|\psi\rangle = 0$$




- Local asymptotic normality (Gaussian approximation):

  In an "uncertainty neighbourhood" of size $n^{-1/2}$ around $\theta_0$, the overlaps of joint states are approximately equal to those of a Gaussian model with QFI = F

$$\left\langle \psi^{\otimes n}_{\theta_0+u/\sqrt{n}} \middle| \psi^{\otimes n}_{\theta_0+v/\sqrt{n}} \right\rangle = \left\langle \psi | e^{i(u-v)G/\sqrt{n}} \middle| \psi \right\rangle^n \longrightarrow e^{(u-v)^2 F/8} = \left\langle \sqrt{F/2}\, u \middle| \sqrt{F/2}\, v \right\rangle$$

- General LAN for mixed states & multi-dimensional models[1]

[1] J. Kahn and MG, Commun. Math. Phys. 2009

# Strong convergence of quantum models

## Definition

Let $\mathcal{Q}_n := \{\rho_{\theta,n} \ : \ \theta \in \Theta\}$ and $\mathcal{Q} := \{\rho_\theta \ : \ \theta \in \Theta\}$.

Then $\mathcal{Q}_n$ converges strongly to $\mathcal{Q}$ if there exist quantum channels $T_n, S_n$ s.t.

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \|T_n(\rho_{\theta,n}) - \rho_\theta\|_1 = 0$$

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \|\rho_{\theta,n} - S_n(\rho_\theta)\|_1 = 0$$

# Strong convergence of quantum models

## Definition

Let $\mathcal{Q}_n := \{\rho_{\theta,n} \ : \ \theta \in \Theta\}$ and $\mathcal{Q} := \{\rho_\theta \ : \ \theta \in \Theta\}$.

Then $\mathcal{Q}_n$ converges strongly to $\mathcal{Q}$ if there exist quantum channels $T_n, S_n$ s.t.

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \|T_n(\rho_{\theta,n}) - \rho_\theta\|_1 = 0$$

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \|\rho_{\theta,n} - S_n(\rho_\theta)\|_1 = 0$$

## Theorem

- If $\rho_{\theta,n} = |\psi_{\theta,n}\rangle\langle\psi_{\theta,n}|$ and $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$ then strong convergence implies

$$\lim_{n\to\infty} \langle\psi_{\theta_1,n} \,|\, \psi_{\theta_2,n}\rangle = \langle\psi_{\theta_1} \,|\, \psi_{\theta_2}\rangle, \qquad \text{( for some choice of phases!)}$$

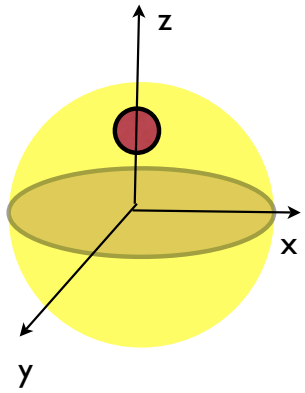
- If $\Theta$ is finite the converse holds as well

# Local spin model and the Gaussian limit

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

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

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

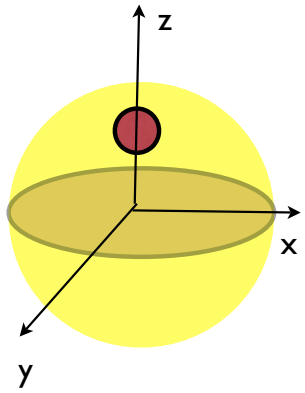


# Local spin model and the Gaussian limit

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

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

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




- Gaussian shift model: $N_{\mathbf{u}} \otimes \Phi_{\mathbf{u}}$
  - Classical part: $N_{\mathbf{u}} := N(u_z, \mu(1-\mu))$
  - Quantum part: $\Phi_{\mathbf{u}} := \Phi\left(u_x\sqrt{2(2\mu-1)}\,,\, u_y\sqrt{2(2\mu-1)}\,;\, (2\mu-1)^{-1}\right)$

# Local asymptotic normality for mixed spin states

## 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 $\eta < 1/4$

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

and

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

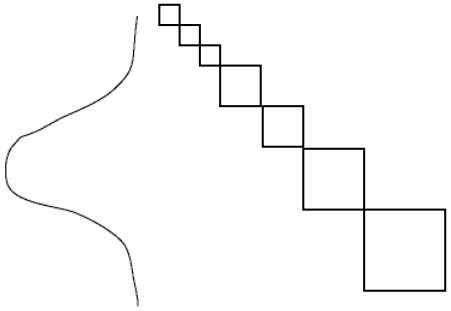
[Guta, Janssens and Kahn, C.M.P. 2008]

# 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 $L^2(\mathbb{R})$

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

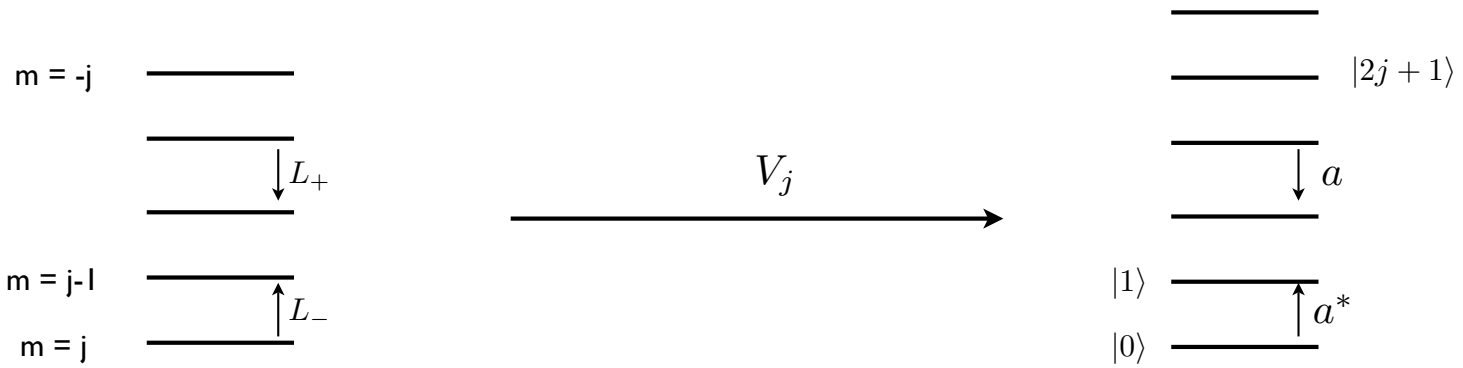
# Isometric embedding

- 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

$$
\left\{ \begin{array}{ll} L_+ & := L_x + iL_y \\ L_- & := L_x - iL_y \end{array} \right. \qquad \text{and} \qquad \left\{ \begin{array}{ll} a & := (Q+iP)/\sqrt{2} \\ a^* & := (Q-iP)/\sqrt{2} \end{array} \right.
$$



# Local asymptotic normality in $d$-dimensions

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

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

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

# 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

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

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

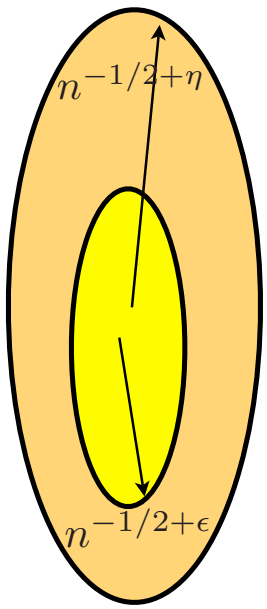
where

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

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

# Asymptotically optimal (adaptive) measurement procedure

Given $n$ i.i.d. spins prepared in state $\rho_\theta$




1. Use $n^{1-\epsilon}$ copies to produce a rough estimator $\rho_0$
2. Map remaining $\tilde{n} = n - n^{1-\epsilon}$ states through $T_{\tilde{n}}$
3. Perform optimal Gaussian measurement and produce estimator

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

# The Holevo bound is achievable

- Holevo bound: quantum statistical model $\{\rho^\theta : \theta \in \Theta \subset \mathbb{R}^k\}$
  - $X_\theta := (X_{\theta,1}, \ldots, X_{\theta,k})$ s.t. $\mathrm{Tr}(\rho^\theta X_{\theta,i}) = 0, \quad \mathrm{Tr}(\frac{\partial \rho^\theta}{\partial \theta_i} X_{\theta,j}) = \delta_{i,j}$
  - $Z(X_\theta)_{i,j} := \mathrm{Tr}(\rho^\theta X_{\theta,j} X_{\theta,i})$

For any unbiased measurement with outcome $\hat{\theta} \in \mathbb{R}^k$

$$\mathbb{E}(\|\hat{\theta} - \theta\|^2) \geq C(\theta) := \inf_{X_\theta} \mathrm{Tr}\left(\mathrm{Re}(Z(X_\theta)) + |\mathrm{Im}(Z(X_\theta))|\right)$$

# The Holevo bound is achievable

- Holevo bound: quantum statistical model $\{\rho^\theta : \theta \in \Theta \subset \mathbb{R}^k\}$
  - $X_\theta := (X_{\theta,1}, \ldots, X_{\theta,k})$ s.t. $\mathrm{Tr}(\rho^\theta X_{\theta,i}) = 0, \quad \mathrm{Tr}(\frac{\partial \rho^\theta}{\partial \theta_i} X_{\theta,j}) = \delta_{i,j}$
  - $Z(X_\theta)_{i,j} := \mathrm{Tr}(\rho^\theta X_{\theta,j} X_{\theta,i})$

For any unbiased measurement with outcome $\hat{\theta} \in \mathbb{R}^k$

$$\mathbb{E}(\|\hat{\theta} - \theta\|^2) \geq C(\theta) := \inf_{X_\theta} \mathrm{Tr}\left(\mathrm{Re}(Z(X_\theta)) + |\mathrm{Im}(Z(X_\theta))|\right)$$

**Theorem** [M.G. and J. Kahn 2010]

There exist measurements $\hat{\theta}_n$ for $\left\{\left(\rho^\theta\right)^{\otimes n} : \theta \in \Theta\right\}$ such that

- the Holevo bound is attained asymptotically

$$\lim_{n\to\infty} n\mathbb{E}(\|\hat{\theta}_n - \theta\|^2) = C(\theta)$$

- asymptotically, $\hat{\theta}_n$ is locally minimax and Bayesian optimal for regular priors

# Blocks indexed by Young diagrams

- Block diagonal form

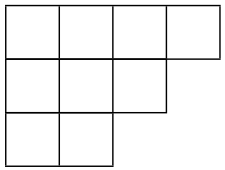
$$\left(\mathbb{C}^d\right)^{\otimes n} = \bigoplus_\lambda \mathcal{H}_\lambda \otimes \mathcal{K}_\lambda$$

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

- Young diagrams $\lambda$ with $d$ lines and $n$ boxes

$\lambda_1 \approx n\mu_1$

$\lambda_d \approx n\mu_d$



- 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

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

# Outlook

- Quantum statistics as non-commutative extension of statistical inference
  - quantum Fisher information, Cramér-Rao bound
  - quantum CLT and local asymptotic normality
  - Quantum Stein Lemma, Chernoff bound
  - ...

- Work in progress
  - Convergence theory of quantum statistical models
  - Non-parametric (infinite dimensional) quantum statistics
  - quantum model selection, efficient estimation of large systems
  - quantum learning
  - ....