

Inferência Bayesiana em Tomografia de Estados Quânticos ¹

Douglas S. Gonçalves

DMA - IMECC - UNICAMP

25 de junho de 2012
Seminários ME 705A

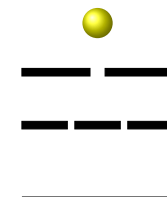
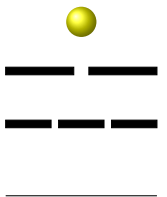
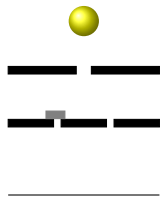
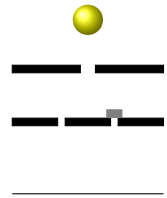
¹Este trabalho conta com o apoio financeiro da FAPESP. 2009/08027-0 

1 Um pouco de física

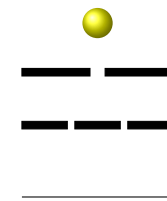
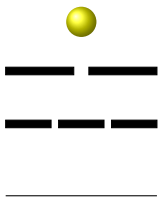
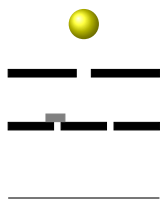
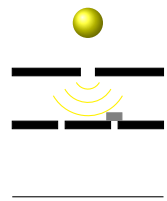
2 Tomografia de Estados Quânticos: Abordagens usuais

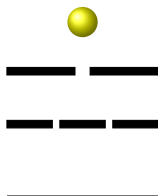
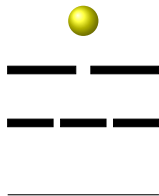
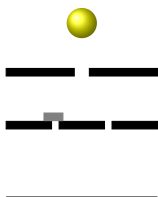
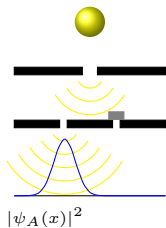
3 Inferência Bayesiana

Dualidade Onda-Partícula



Dualidade Onda-Partícula



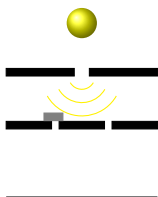
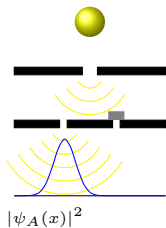


$\psi(x)$ função de onda

$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

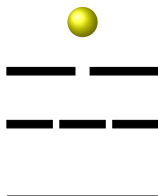
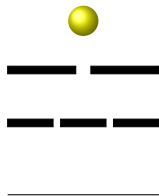


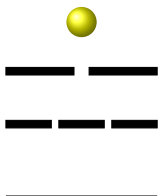
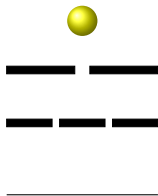
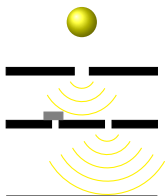
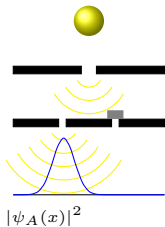
$\psi(x)$ função de onda

$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$



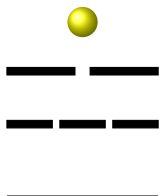
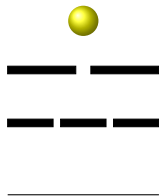
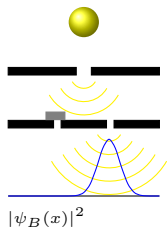
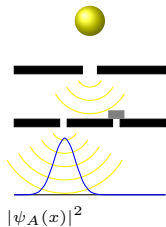


$\psi(x)$ função de onda

$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

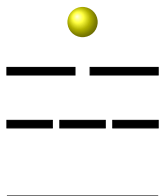
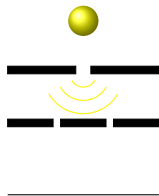
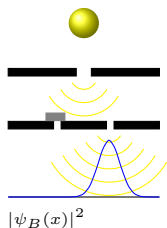
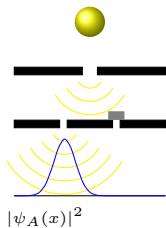


$\psi(x)$ função de onda

$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$



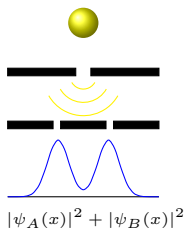
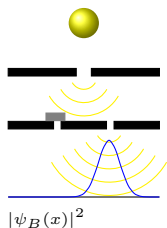
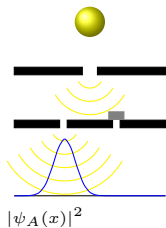
$\psi(x)$ função de onda

$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

A e B independentes



$\psi(x)$ função de onda

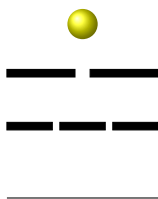
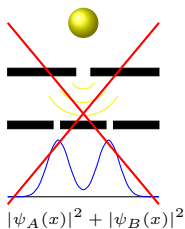
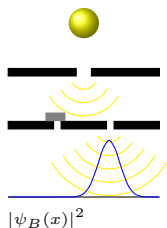
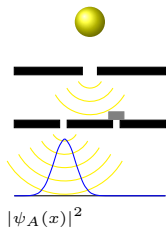
$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

A e B independentes

$$P(A \cup B) = |\psi_A(x)|^2 + |\psi_B(x)|^2$$



$\psi(x)$ função de onda

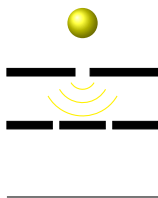
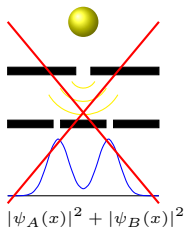
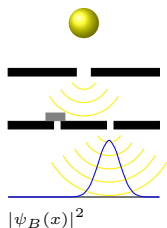
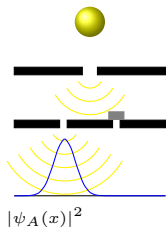
$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

A e B independentes

$$P(A \cup B) \neq |\psi_A(x)|^2 + |\psi_B(x)|^2$$



$\psi(x)$ função de onda

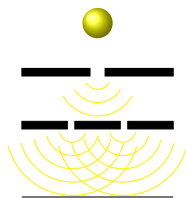
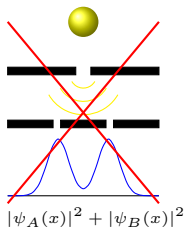
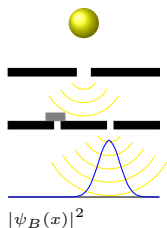
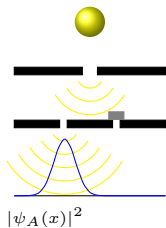
$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

A e B independentes

$$P(A \cup B) \neq |\psi_A(x)|^2 + |\psi_B(x)|^2$$



$\psi(x)$ função de onda

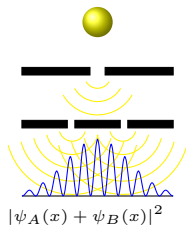
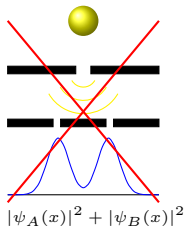
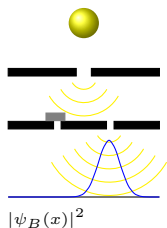
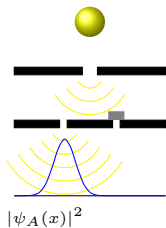
$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

A e B independentes

$$P(A \cup B) \neq |\psi_A(x)|^2 + |\psi_B(x)|^2$$



$\psi(x)$ função de onda

$|\psi(x)|^2$ probabilidade de encontrar uma partícula em x em um dado tempo

$$P(A) = |\psi_A(x)|^2$$

$$P(B) = |\psi_B(x)|^2$$

A e B independentes

$$P(A \cup B) \neq |\psi_A(x)|^2 + |\psi_B(x)|^2$$

$$P(A \cup B) = |\psi_A(x) + \psi_B(x)|^2$$

Clássica	Quântica
Ω	\mathcal{H}
\mathcal{F}	\mathcal{L}
$\Pr_c : \mathcal{F} \rightarrow [0, 1]$	$\Pr_q : \mathcal{L} \rightarrow [0, 1]$

Axiomas

(C1) $\Pr_c(\emptyset) = 0$

(C2) $\Pr_c(A^c) = 1 - \Pr_c(A)$

(C3) $\Pr_c(A \cup B) = \Pr_c(A) + \Pr_c(B) - \Pr_c(A \cap B)$

(Q1) $\Pr_q(\mathbf{0}) = 0$

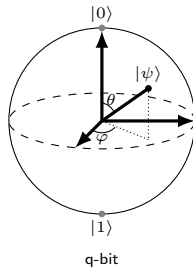
(Q2) $\Pr_q(V^\perp) = 1 - \Pr_q(V)$

(Q3) $\Pr_q(V \oplus W) = \Pr_q(V) + \Pr_q(W), \quad V \perp W$

Variáveis discretas: propriedades de sistemas quânticos descritas por um número finito de eventos

Exemplos: níveis de energia, spins, polarização

Q-bit: sistema quântico de dois níveis



$|\psi\rangle$ vetor unitário no espaço de Hilbert \mathcal{H}

$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ (**Superposição de estados**)

Ensemble de estados quânticos $\{|\psi_i\rangle\}_i$

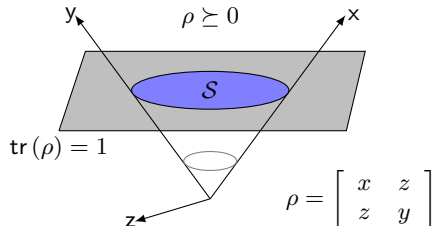
$$\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|,$$

com $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$.

ρ é chamada **matriz de densidade**.

Propriedades

- $\rho \succeq 0$, ie, $\langle\phi|\rho|\phi\rangle \geq 0, \forall|\phi\rangle$
- $\text{tr}(\rho) = 1$



Toda medição sobre um sistema quântico em um estado ρ é descrita por um conjunto de operadores $\{E_i\}_i$ tais que

- $E_i \succeq 0$
- $\sum_i E_i = I$

onde i diz respeito a cada um dos possíveis resultados da medição.

A probabilidade de ocorrência do evento i é dada pela **Regra de Born**

$$p_i = \text{tr}(E_i \rho).$$

Daí o nome matriz de densidade. Assim como

$$\langle X \rangle = \sum_i x_i p(x_i),$$

temos que

$$\langle O \rangle = \text{tr}(\rho O) = \text{tr} \left(\rho \left(\sum_i \lambda_i \Pi_i \right) \right) = \sum_i \lambda_i \text{tr}(\Pi_i \rho).$$

Toda medição sobre um sistema quântico em um estado ρ é descrita por um conjunto de operadores $\{E_i\}_i$ tais que

- $E_i \succeq 0$
- $\sum_i E_i = I$

onde i diz respeito a cada um dos possíveis resultados da medição.

A probabilidade de ocorrência do evento i é dada pela **Regra de Born**

$$p_i = \text{tr}(E_i \rho).$$



Inversão Linear

- $\rho \in \mathbb{C}^{d \times d}$
- d^2 operadores E_i (d^2 equações)

- Teoria:

$$f_i \approx p_i = \text{tr}(E_i \rho), \quad \forall i$$

- Sistema linear: fornece uma estimativa para ρ

- Prática:

$$f_i = \varepsilon_i + p_i = \text{tr}(E_i \rho), \quad \forall i$$

- Sistema linear: leva a estimativas “fora do espaço paramétrico” $\rho \notin \mathcal{S}$

Uma alternativa seria

$$\begin{aligned} \min_{\rho} \quad & \sum_i (\text{tr}(E_i \rho) - f_i)^2 \\ \text{s.a} \quad & \text{tr}(\rho) = 1 \\ & \rho \succeq 0 \end{aligned}$$

quadrados mínimos restrito.

Ou, assumindo algum modelo paramétrico

$$\begin{aligned} \max_{\rho} \quad & \mathcal{L}(\rho) \equiv P(n | \rho) \\ \text{s.a} \quad & \text{tr}(\rho) = 1 \\ & \rho \succeq 0, \end{aligned}$$

temos a estimativa de **Máxima Verossimilhança**.

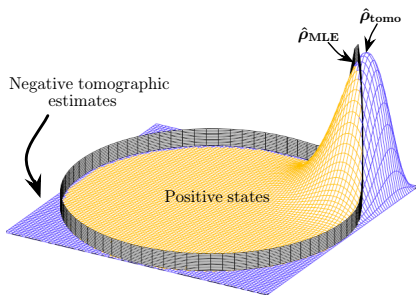
- Multinomial (alta eficiência, RMN, ion trap)

$$P(n | \rho) = \frac{N!}{n_1! \dots n_m!} \prod_{i=1} \text{tr}(E_i \rho)^{n_i}$$

- Gaussiana (photon detection)

$$P(n | \rho) = \frac{1}{N_0} \prod_i \exp - \frac{1}{2} \frac{(N \text{tr}(E_i \rho) - n_i)^2}{N \text{tr}(E_i \rho)}$$

- Autovalores nulos (**probabilidades nulas**)
- Incompatível com “barras de erro”



R. Blume-Kohout, New J. of Physics, 2010

- Visão frequentista × Visão Bayesiana
- **Frequentista:** probabilidade como limite de frequências
- **Bayesiana:** probabilidade como medida de credibilidade
- O conhecimento prévio é atualizado pelos dados
- O parâmetro a ser estimado é considerado uma variável aleatória e definido pela distribuição a posteriori

Inferência Bayesiana

$$P(H | D) \propto P(D | H)P(H),$$

em Tomografia de Estados Quânticos

$$\pi(\rho) \propto \mathcal{L}(\rho)\pi_0(\rho).$$

Estimativa Bayesiana (média)

$$\rho_B = \int_{\mathcal{S}} \rho \pi(\rho) d\rho,$$

lembrando que $\mathcal{S} = \{\rho = \rho^\dagger \mid \text{tr}(\rho) = 1, \rho \succeq 0\}$.

Verossimilhança Multinomial

$$\mathcal{L}(\rho) = P(n \mid \rho) = \frac{N!}{n_1! \dots n_m!} \prod_{i=1}^m \text{tr}(E_i \rho)^{n_i}.$$

Priori não-informativa: $\pi_0(\rho)$ uniforme em \mathcal{S}

Garantia de estimativas de **posto completo**

Como avaliar

$$\int_{\mathcal{S}} \rho \pi(\rho) d\rho,$$

ou amostrar de $\pi(\rho)$??

Monte Carlo via Cadeias de Markov (MCMC)

Amostrar de uma cadeia de Markov $\{\rho_t\}_t$ que tenha $\pi(\cdot)$ como distribuição estacionária

Após o burn-in, tomamos a média ergódica

$$\rho_B = \int_S \rho \pi(\rho) d\rho \approx \frac{1}{n-m} \sum_{t=m}^n \rho_t$$

Metropolis-Hastings

Candidato Y amostrado de $q(\cdot | \rho_t)$, aceito quando $u < \alpha(\rho_t, Y)$, $u \sim U(0, 1)$

$$\alpha(\rho_t, Y) = \min \left(1, \frac{\pi(Y)q(\rho_t | Y)}{\pi(\rho_t)q(Y | \rho_t)} \right)$$

A distribuição proposta $q(\cdot | \rho_t)$

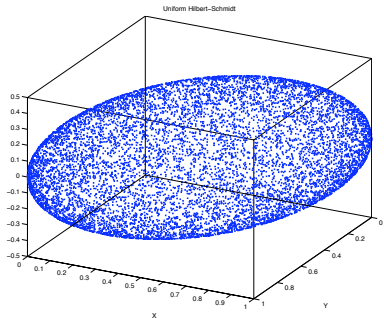
$$Y \sim (1 - \lambda)\rho_t + \lambda X,$$

com $\lambda \in (0, 1)$ e X de uma distribuição uniforme em \mathcal{S} (medida de Hilbert-Schmidt).

Primeiro tomamos W do ensemble de Ginibre: W_{ij} normais padrão i.i.d., e então

$$X = \frac{W^\dagger W}{\text{tr}(W^\dagger W)}$$

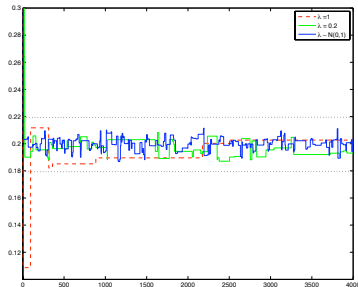
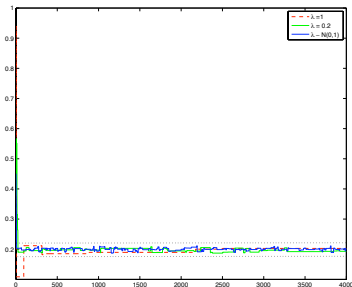
Pela **convexidade** de \mathcal{S} , para $\lambda \in (0, 1)$ temos que $(1 - \lambda)\rho_t + \lambda X \in \mathcal{S}$.



Escolha de $\lambda \in (0, 1)$:

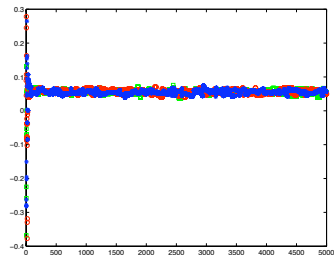
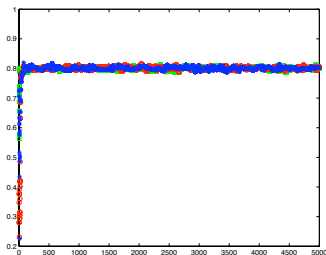
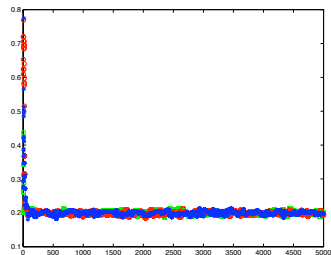
$$Y \sim (1 - \lambda)\rho_t + \lambda X,$$

- $\lambda = 1$, distribuição uniforme em \mathcal{S}
- $\lambda = \theta$, fixo em $(0, 1)$, distribuição uniforme em $\mathcal{A} \subset \mathcal{S}$ (suporte variável)
- $\lambda \sim N(0, \sigma^2)$, favorecer valores de λ mais próximos de zero. ($\sigma = 0.3$)

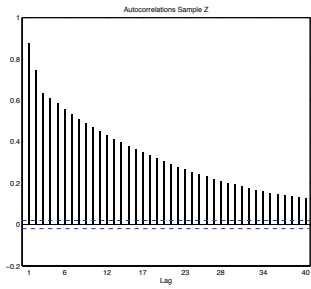
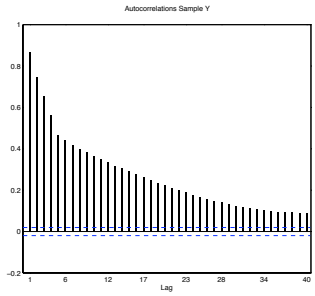
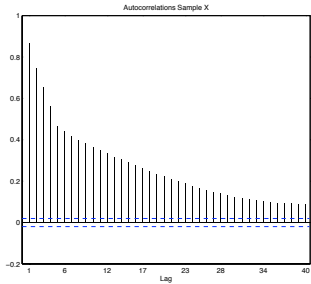


Burn-in = 1000

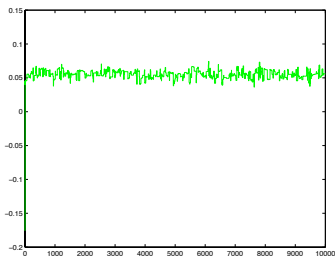
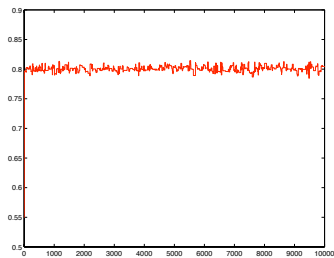
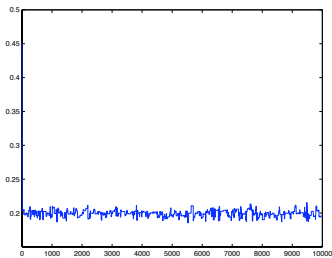
(Cadeias se misturam rapidamente)



Autocorrelações



Inferência Bayesiana na Tomografia de Estados Quânticos



$$\rho_* = \begin{bmatrix} 0.2 & 0.05 \\ 0.05 & 0.8 \end{bmatrix}$$

$$\rho_B = \begin{bmatrix} 0.1993 & 0.0549 \\ 0.0549 & 0.8007 \end{bmatrix}$$

Diagnósticos de convergência

- Gelman-Rubin

$m \geq 2$ cadeias, iniciadas em pontos “espalhados”

Calcular a variância “dentre” as cadeias W e “entre” as cadeias B

$$\hat{R} = \sqrt{\frac{\text{Var}(\theta)}{W}}, \quad \text{onde} \quad \text{Var}(\theta) = \left(1 - \frac{1}{n}\right)W + \frac{1}{n}B$$

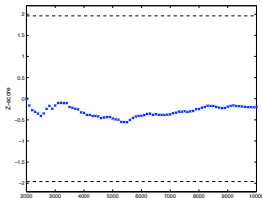
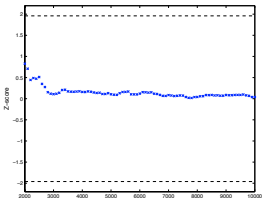
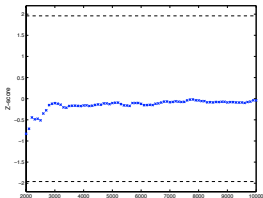
$\hat{R} < 1.1$ indicativo de convergência

- Geweke

Uma única cadeia

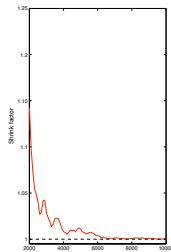
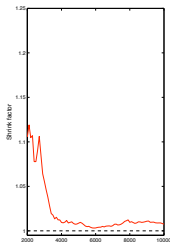
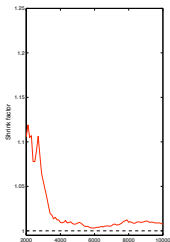
Teste de igualdade de médias: primeiros 10% e últimos 50% (após burn-in)

■ Geweke plots



■ Gelman-Rubin

3 cadeias em paralelo



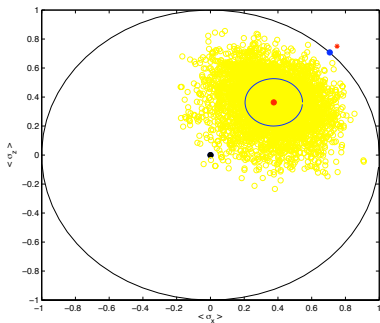
Exemplo (maldoso): medir os projetores $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$

$|0\rangle\langle 0| : 14, |1\rangle\langle 1| : 2, |+\rangle\langle +| : 14, |-\rangle\langle -| : 2, \langle \sigma_y \rangle = 0$

$$\begin{aligned}\rho_{IL} &= \frac{1}{2} \left(\mathbb{I} + \frac{3}{4} \sigma_x + \frac{3}{4} \sigma_z \right) \\ &= \begin{bmatrix} 0.8750 & 0.3750 \\ 0.3750 & 0.1250 \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\rho_{ML} &= \frac{1}{2} \left(\mathbb{I} + \frac{1}{\sqrt{2}} \sigma_x + \frac{1}{\sqrt{2}} \sigma_z \right) \\ &= \begin{bmatrix} 0.8536 & 0.3536 \\ 0.3536 & 0.1464 \end{bmatrix}\end{aligned}$$

$$\rho_B = \begin{bmatrix} 0.6817 & 0.1876 \\ 0.1876 & 0.3183 \end{bmatrix}$$



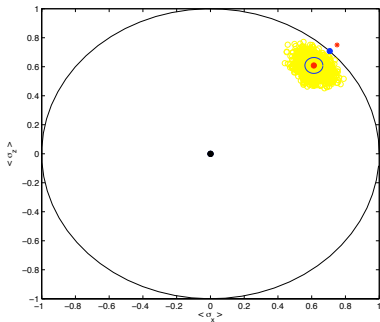
Exemplo (maldoso): medir os projetores $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$

$|0\rangle\langle 0| : 14, |1\rangle\langle 1| : 2, |+\rangle\langle +| : 14, |-\rangle\langle -| : 2, \langle \sigma_y \rangle = 0$

$$\begin{aligned}\rho_{IL} &= \frac{1}{2} \left(\mathbb{I} + \frac{3}{4} \sigma_x + \frac{3}{4} \sigma_z \right) \\ &= \begin{bmatrix} 0.8750 & 0.3750 \\ 0.3750 & 0.1250 \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\rho_{ML} &= \frac{1}{2} \left(\mathbb{I} + \frac{1}{\sqrt{2}} \sigma_x + \frac{1}{\sqrt{2}} \sigma_z \right) \\ &= \begin{bmatrix} 0.8536 & 0.3536 \\ 0.3536 & 0.1464 \end{bmatrix}\end{aligned}$$

$$\rho_B = \begin{bmatrix} 0.8044 & 0.3065 \\ 0.3065 & 0.1956 \end{bmatrix}$$



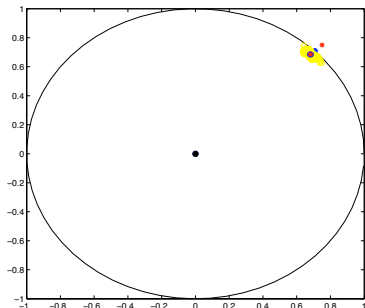
Exemplo (maldoso): medir os projetores $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$

$|0\rangle\langle 0| : 14, |1\rangle\langle 1| : 2, |+\rangle\langle +| : 14, |-\rangle\langle -| : 2, \langle \sigma_y \rangle = 0$

$$\begin{aligned}\rho_{IL} &= \frac{1}{2} \left(\mathbb{I} + \frac{3}{4}\sigma_x + \frac{3}{4}\sigma_z \right) \\ &= \begin{bmatrix} 0.8750 & 0.3750 \\ 0.3750 & 0.1250 \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\rho_{ML} &= \frac{1}{2} \left(\mathbb{I} + \frac{1}{\sqrt{2}}\sigma_x + \frac{1}{\sqrt{2}}\sigma_z \right) \\ &= \begin{bmatrix} 0.8536 & 0.3536 \\ 0.3536 & 0.1464 \end{bmatrix}\end{aligned}$$

$$\rho_B = \begin{bmatrix} 0.8425 & 0.3407 \\ 0.3407 & 0.1575 \end{bmatrix}$$



Comentários

- É possível aplicar IB em TEQ
- Dependendo da priori temos garantia de estimativas de posto completo
- Compatibilidade com “barras de erro”

Trabalhos futuros

- Uso de prioris informativas
- Generalização da distribuição de Dirichlet
- Testar MCMC em sistemas de dimensão maior
- Eficiência dos algoritmos MCMC