Shot Noise

Basics - Estimators

$|\langle \lambda_m | \lambda_{m'} \rangle|^2 = \delta_{mm'}$

generally assume these are real numbers but not essential

$$M = \sum_{m=1}^{d} \lambda_m |\lambda_m\rangle\langle\lambda_m| \qquad \text{Observable}$$

Want to estimate $\bar{M} = \mathrm{Tr}(\rho M)$

Strategy - measure ρ in eigenbasis of M & average

Estimator:

$$\hat{M}_N = \frac{1}{N} \sum_{s=1}^{N} \lambda_{m(s)}$$

eigenvalue of M obtained on sth shot measuring ρ

$$\mathbb{E}(\hat{M}_N) = \frac{1}{N} \sum_{s=1}^{N} \underbrace{\mathbb{E}(\lambda_{m(s)})}_{\underbrace{\sum_{m(s)} \lambda_{m(s)}\, \rho(\lambda_{m(s)})}_{\mathrm{Tr}(M\rho)}}$$

$\langle \lambda_{m(s)} | \rho | \lambda_{m(s)} \rangle = \mathrm{Tr}(\rho\, |\lambda_{m(s)}\rangle\langle\lambda_{m(s)}|)$

$$\Rightarrow \underbrace{\mathbb{E}(\hat{M}_N) = \frac{1}{N} \sum_{s=1}^{N} \mathrm{Tr}(M\rho) = \mathrm{Tr}(M\rho) = \bar{M}}$$

ie. The average of the estimator for finite N is equal to the true average

"$\hat{M}_N$ is an unbiased estimator"

$$\mathrm{Var}(\hat{M}_N) = \mathcal{E}(\hat{M}_N^2) - \underbrace{\mathcal{E}(\hat{M}_N)^2}_{\mathrm{Tr}(\rho M)^2}$$

$$\mathcal{E}(\hat{M}_N^2) = \mathcal{E}\left(\frac{1}{N^2}\sum_{s,t=1}^{N} \lambda_{m(s)}\lambda_{m(t)}\right)$$

$$= \frac{1}{N^2}\left(\sum_{s=1}^{N} \underbrace{\mathcal{E}(\lambda_{m(s)}^2)}_{\sum_m \lambda_{m(s)}^2 P(\lambda_{m(s)})} + \sum_{s\neq t=1}^{N} \underbrace{\mathcal{E}(\lambda_{m(s)}\lambda_{m(t)})}_{\mathcal{E}(\lambda_{m(s)})\mathcal{E}(\lambda_{m(t)})}\right)$$

$$\sum_m \lambda_{m(s)}^2 P(\lambda_{m(s)}) = \sum_m \lambda_{m(s)}^2 \mathrm{Tr}(\rho\, |\lambda_{m(s)}\rangle\langle\lambda_{m(s)}|) = \mathrm{Tr}(\rho M^2)$$

$$\mathcal{E}(\lambda_{m(s)})\mathcal{E}(\lambda_{m(t)}) = \mathrm{Tr}(M\rho)^2$$

$$= \frac{1}{N^2}\left(N\,\mathrm{Tr}(M^2\rho) + (N^2-N)\,\mathrm{Tr}(M\rho)^2\right)$$

$$= \frac{1}{\cancel{N^2}}\left(\cancel{N^2}\,\mathrm{Tr}(M\rho)^2 + \underbrace{\cancel{N}}_{1/N}\left(\mathrm{Tr}(M^2\rho) - \mathrm{Tr}(M\rho)^2\right)\right)$$

$$\Rightarrow \underbrace{\mathrm{Var}(\hat{M}_N)}_{\text{Estimate precision } \epsilon} = \frac{1}{N}\Big(\underbrace{\mathrm{Tr}(M^2\rho) - \mathrm{Tr}(M\rho)^2}_{\text{Standard quantum variance} \,=\, \mathrm{Var}_\rho(M)}\Big)$$

$$\epsilon = \frac{\sqrt{\mathrm{Var}_\rho(M)}}{\sqrt{N}} \equiv \underbrace{N = \frac{\mathrm{Var}_\rho(M)}{\epsilon^2}}_{\text{Shots required to estimate } \mathrm{Tr}(M\rho) \text{ to precision } \epsilon}$$

Examples

① Consider two different ways of computing the fidelity between two states

A. Loschmidt Echo circuit

$$|\langle\psi|\phi\rangle|^2 = \langle\psi|\overbrace{R_\phi|0\rangle}^{|\phi\rangle}\langle 0|R_\phi^\dagger|\psi\rangle$$

$$= \langle\psi|\, R_\phi\, M_A\, R_\phi^\dagger\,|\psi\rangle$$

$M_A = |0\rangle\langle 0|$

B. SWAP Test

$$|\langle\psi|\phi\rangle|^2 = \langle\psi|\otimes\langle\phi|\ \overbrace{\text{SWAP}}^{M_B}\ |\psi\rangle\otimes|\phi\rangle$$

$$= \sum_{ij,\,i'j'} \psi_i^* \phi_j^* \psi_{i'} \phi_{j'}\ \underbrace{\langle ij|\underbrace{\text{SWAP}|i'j'\rangle}_{|j'i'\rangle}}_{\delta_{ij'}\delta_{i'j}}$$

$$= \sum_{ij} \psi_i^* \psi_j \phi_j^* \phi_i$$

$$= \left(\sum_i \psi_i^* \phi_i\right)\left(\sum_i \psi_i^* \phi_i\right)^*$$

$$= |\langle\psi|\phi\rangle|^2$$

Check that this does actually work

Which converges quicker? i.e. which exhibits less shot noise

$$\text{Var}(\hat{M}_A(N)) = \frac{\text{Var}_{|\psi\rangle}(M_A)}{N} = \frac{\langle\psi|R_\phi^\dagger \overbrace{(|0\rangle\langle 0|)^2}^{|0\rangle\langle 0|} R_\phi|\psi\rangle - |\langle\psi|\phi\rangle|^4}{N}$$

$$= \frac{|\langle\psi|\phi\rangle|^2 - |\langle\psi|\phi\rangle|^4}{N}$$

$$\text{Var}(\hat{M}_B(N)) = \frac{\text{Var}_{|\psi\rangle}(M_B)}{N} = \frac{\langle\psi|\langle\phi|\ \overbrace{\text{SWAP}^2}^{I}\ |\psi\rangle|\phi\rangle - |\langle\psi|\phi\rangle|^4}{N}$$

$$= \frac{1 - |\langle\psi|\phi\rangle|^4}{N}$$

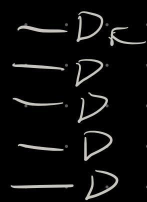
As $|\langle\psi|\phi\rangle|^2 \leq 1$

Shot noise is lower for method A!

② Marco's Problem:

Two different ways of measuring $\langle\psi|\sigma_z^{\otimes n}|\psi\rangle$

Ⓐ ρ

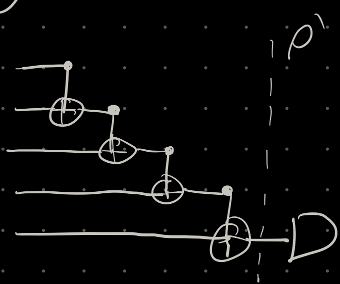


computational basis measurement

$$\underbrace{\langle Z^{\otimes n}\rangle_\rho}$$

$= 1$ if even no. of 1s
$= -1$ if odd " " "

Ⓑ ρ



$$\underbrace{\langle I \otimes \sigma_z\rangle_{\rho'}}$$

$\rho' = |i_0 \oplus i_1 \oplus \dots \oplus i_n\rangle$

$\begin{cases} = |0\rangle & \text{if even no. of 1s} \\ = |1\rangle & \text{if odd "} \end{cases}$

$= 1$ if even no. of 1s
$= -1$ if odd " " "

agree!

So which is better?

$M_A = Z^{\otimes n}$ & ρ
$M_B = I \otimes Z$ & ρ'

$\mathrm{Tr}(M_A^2 \rho) = \mathrm{Tr}(I\rho) = 1$
$\mathrm{Tr}(M_B^2 \rho) = \mathrm{Tr}(I\rho') = 1$

$\Rightarrow$ Variances are the same.

Probability Bounds

Core question:

"Given a random variable Z with expectation $E(Z)$ - how likely is Z to be close to its expectation $E(Z)$?"

Or, in the context of computing quantum averages:

" Given a quantum state ρ & observable M what is the probability that the estimator M_N is close to its average $Tr(M\rho)$? "

Markov's Inequality Let $Z \geq 0$

$$P(Z \geq t) \leq \frac{E(Z)}{t}$$

Proof*:

* probably wouldn't satisfy a mathematician

$$E(Z) = E(Z \mid Z \geq t)\, P(Z \geq t) + \underbrace{E(Z \mid Z < t)}_{\substack{+ve \\ \text{as } Z \geq 0}}\, \underbrace{P(Z < t)}_{+ve}$$

$$\geq \underbrace{E(Z \mid Z \geq t)}_{\geq t}\, P(Z \geq t)$$

$$\Rightarrow E(Z) \geq t\, P(Z \geq t)$$

$$\Rightarrow \quad P(Z \geq t) \leq \frac{E(Z)}{t}$$

Chebyshev's Inequality

Super useful bound!

Let Z be a random variable with $Var(Z) \leq \infty$

$$P(|Z - E(Z)| \geq t) \leq \frac{Var(Z)}{t^2}$$

(for $t > 0$)

Proof:

$$P(|Z - E(Z)| \geq t) = P((Z - E(Z))^2 \geq t^2)$$

Markov

$$\leq \frac{E((Z - E(Z))^2)}{t^2}$$

$$= \frac{Var(Z)}{t^2}$$

Consequences:

1. Provides a bound on how quickly estimators converge

$$P(|\hat{M}_N - Tr(M\rho)| \geq t) \leq \frac{Var(\hat{M}_N)}{t^2} = \frac{Var_\rho(M)}{N t^2}$$

eg. given an error probability tolerance δ we need

$$N \geq \frac{\mathrm{Var}_\rho(M)}{\delta t^2} \quad \text{shots}$$

(2) Central to studies of exponential concentration

Probabilistic exponential concentration

$Z(\alpha)$ is probabilistically exponentially concentrated if

$$P_\alpha(|Z(\alpha) - \mu| \geq t) \leq \frac{\beta}{t^2}$$

with $\beta \in \mathcal{O}(1/b^n)$ for $b > 1$

eg. $Z(\alpha) = C(\theta) = \mathrm{Tr}(H\, U(\theta)\, \rho\, U(\theta)^\dagger)$

Cost function for variational quantum algorithm

parameterized circuit

We can use Chebyshev's Inequality to study cost landscape

$$P_\theta(|C(\theta) - \mu| \geq t) \leq \frac{\mathrm{Var}_\theta(C(\theta))}{t^2}$$

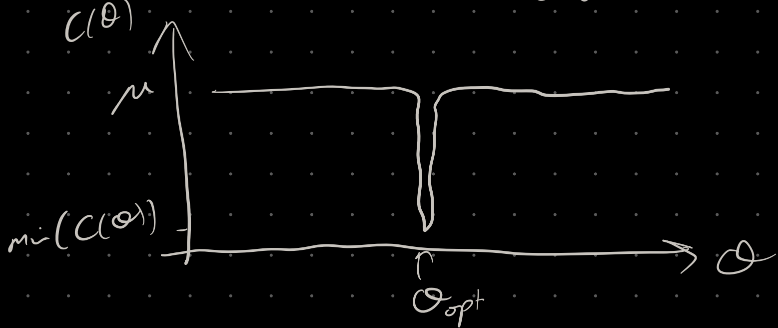
key thing to compute

If $\mathrm{Var}_\theta(C(\theta)) \sim 1/2^n$

Then $P_\theta(|C(\theta)-\mu| \geq t) \leq \dfrac{1}{2^n t^2}$

$\Rightarrow$ Cost is exponentially concentrated

$\Rightarrow$ landscape is very flat



Hoeffding's Inequality

$$P(|S_N - E(S_N)| \geq t) \leq 2e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}$$

$S_N = \sum_{i=1}^{N} X_i$ ← Independent Random variables s.t. $a_i \leq X_i \leq b_i$

(For proof see typed notes)

I've intensionally used the statement of Hoeffding's from wikipedia here because that's where I always go when I forget the inequality and then it takes a couple of lines to translate it into a more useful form.

Applied to measuring M for system in state ρ

$$X_i \to \frac{\lambda_{m(i)}}{N} \qquad S_N = \frac{1}{N}\sum_{s=1}^{N} \lambda_{m(s)} = \hat{M}_N$$

$$\forall s : \quad \frac{\lambda_{min}}{N} \leq \frac{\lambda_{m(s)}}{N} \leq \frac{\lambda_{max}}{N}$$

$$\Rightarrow \sum_{i=1}^{N} (b_i - a_i) = \sum_{i=1}^{N} \frac{(\lambda_{max} - \lambda_{min})^2}{N^2} = \frac{(\lambda_{max} - \lambda_{min})^2}{N}$$

Therefore Hoeffding's can be more usefully restated as

$$\boxed{P(|\hat{M}_N - M| \geq \epsilon) \leq 2e^{-2N\epsilon^2/(\lambda_{max} - \lambda_{min})^2}}$$

Or, equivalently

$\Delta := \lambda_{max} - \lambda_{min}$

let $\delta/2 = e^{-2N\epsilon^2/\Delta^2}$

probability of error tolerance

$$\log(\delta/2) = -\frac{2N\epsilon^2}{\Delta^2}$$

$$N = -\frac{\Delta^2}{2\epsilon^2}\log(\delta/2) = \frac{\Delta^2}{2\epsilon^2}\log(2/\delta)$$

Need $N \geq \frac{\Delta^2}{2\epsilon^2}\log(2/\delta)$ to fail with $P < \delta$

Compare with equivalent bound from Chebyshev $\quad N \geq \frac{Var_p(M)}{\delta\epsilon^2}$

$\log(2/\delta) < 1/\delta$ but $Var_p(M) \leq \Delta^2$

$Var_p(M) \leq$ or $\geq \frac{\Delta^2}{2}$

$\Rightarrow$ Which is tighter varies!

Revisiting examples.

1. Eigenvalues of $|0\rangle\langle 0|$ are $0, 1 \Rightarrow \Delta = 1$

   $SWAP = \begin{pmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{pmatrix}$ are $\underbrace{1, -1}_{\Rightarrow \Delta = 2}$

   Convergence via Loschmidt Echo is faster than via SWAP test.

2. $\sigma_z^{\otimes n}$ & $I \otimes \sigma_z$ both have eigenvalues $1, -1 \Rightarrow \Delta = 2$

   & convergence rate is the same.