In quantum state tomography, one often makes statements based on the *likelihood function* $\mathcal{L}(\rho)$. The function is defined for a given dataset obtained during the tomography procedure, and is a function on state space. The likelihood is defined as the probability that from the state ρ, the quantum measurements you performed would give you the actual dataset that you observed,

$$\mathcal{L}(\rho) = \Pr[\text{observed data} \,|\, \rho] \ . \tag{1}$$

It is often convenient for various applications to consider the logarithm of this function, the loglikelihood $\ell(\rho) = \ln \mathcal{L}(\rho)$.

More generally, the likelihood and loglikelihood are basic quantities for parameter estimation in statistics (not only in quantum mechanics).

One simple tomography procedure, known as *maximum likelihood estimation*, is to take measurements, write down the likelihood function, and report the maximum of this function as being a suitable candidate for an estimation ρ_{MLE} of what the quantum state of our system was. (Technically, one usually prefers for large amount of data to numerically maximize the loglikelihood, since the latter is concave.)

Exercise 1. ***State Tomography of a Coin.***

Suppose that you have a biased coin, that gives "heads" with probability p and gives "tails" with probability $1-p$. You would like to estimate this bias, based on a finite number of coin flips.

Suppose that you flip the coin N times, and obtain f_h times heads and $f_t = N - f_h$ times tails.

(a) Write down the likelihood function $\mathcal{L}(p)$ and the loglikelihood function $\ell(p)$ for these measurements, as a function of f_h and f_t. Plot these functions in the interval $p \in [0,1]$.

Solution. The likelihood function is given as follows. Recall the definition

$$\mathcal{L}(p) = \Pr[\text{observed data} \,|\, p] \ . \tag{S.1}$$

The probability involved is the probability of observing a sequence of N coin flips with f_h times heads and f_t times tails, i.e.

$$\mathcal{L}(p) = p^{f_h} \times (1-p)^{f_t} \ . \tag{S.2}$$

If one takes the logarithm of $\mathcal{L}(p)$, the multiplications become a sum and $\ell(p)$ is simply given by

$$\ell(p) = f_h \ln(p) + f_t \ln(1-p) \ . \tag{S.3}$$

The plots of these functions for 50 coin flips which yielded 15 heads outcomes and 35 tails outcomes are given in figure 1. Notice how the likelihood peaks around the value 0.3, but with extremely low values ($\sim 10^{-14}$), and how the loglikelihood is much better (at least numerically) well-behaved.

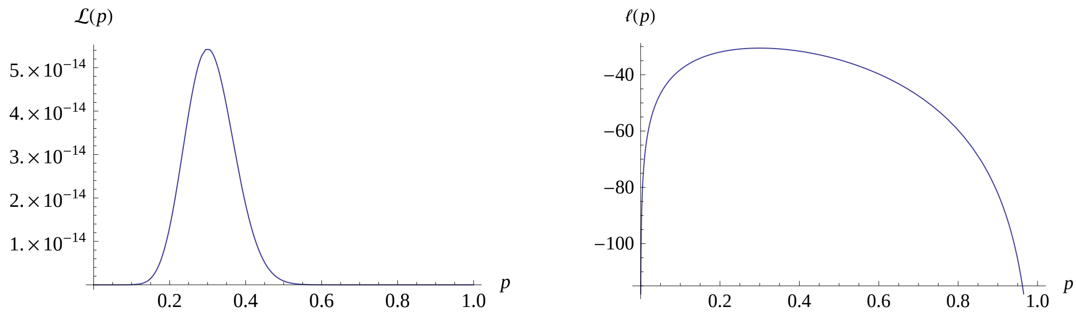


Figure 1: Likelihood and loglikelihood for 50 coin flips with 15 heads outcomes and 35 tails outcomes.

(b) What would you report as the "true" bias of the coin?

Solution. The first (naive) solution is to estimate the probability of getting heads based on how many times the heads outcome actually came out, i.e. $f_h/(f_h+f_t)$. In the example presented in Fig. 1, this corresponds to a bias of $p = 0.3$. Also, by looking at the graph of the likelihood function, we already feel that we should report a bias around 0.3, which seems to be the "most likely". This is motivated by the fact that we expect "not to win at the lottery", i.e. we don't expect events with very low probability to come out.

We can formally calculate the maximum likelihood estimate p_{MLE} by cancelling the derivative $(d\ell)/(dp)$: using (S.3),

$$0 = \frac{d\ell}{dp} = \frac{f_h}{p} - \frac{f_t}{1-p} \quad \Rightarrow \quad 0 = f_h\,(1-p) - f_t\,(p) = f_h - p\,(f_h + f_t) \ , \tag{S.4}$$

which, solved for p, yields $p = f_h/\,(f_h + f_t)$. This is the same as our naive guess of approximating the probability based on the frequencies (which is in fact linear inversion).

This procedure, still, has some intrinsic drawbacks: if, for example, one flips a fair coin only once, then one of the outcomes "heads" or "tails" will come out, i.e. $f_h = 1$ and $f_t = 0$ or the other way around. In this case, our procedure will report $p = 0$ or $p = 1$ but not $p = 1/2$ (!). This is due to the very low number of measurements and to the lack of a proper error bar on the estimation of p. This problem is addressed by reporting confidence regions instead of a single point, for example.

Remark. This problem is identical to measuring only σ_z on an unknown qubit, and estimating the Z-coordinate of its state on the Bloch sphere.

Exercise 2. ***Quantum State Tomography on a Qubit.***

Suppose you have a qubit on which you perform tomographic measurements in order to determine its state ρ_{true}. You measure in the Pauli basis, performing n_x measurements of the σ_x observable, n_y of the σ_y observable, and n_z for σ_z. You accumulate statistics by keeping track of the number f_i^+ of "+1"'s measured for the observable σ_i, and the number f_i^- of "−1"'s observed for that observable.

(a) Write down the likelihood function and the loglikelihood function for this procedure, as a function of the f_i^s's. Show that the loglikelihood function is concave.

Solution. The likelihood is in general given, we recall, as

$$\mathcal{L}\,(\rho) = \Pr\left[\text{observed data}\,|\,\rho\right] \ . \tag{S.5}$$

Here, the probability is given to us by elementary quantum mechanical measurements: the probability of measuring e.g. the outcome "+" when measuring observable σ_i on a quantum system in state ρ is given by $\operatorname{tr}\big(\Pi_i^+\rho\big)$, where Π_i^s is the projector onto the s eigenspace of σ_i. The sequence of measurements that we actually obtained (characterized by the f_i^s) would have occured from ρ with probability

$$\mathcal{L}\,(\rho) = \big(\operatorname{tr}\Pi_x^+\rho\big)^{f_x^+} \times \big(\operatorname{tr}\Pi_x^-\rho\big)^{f_x^-} \times \big(\operatorname{tr}\Pi_y^+\rho\big)^{f_y^+} \times \cdots \times \big(\operatorname{tr}\Pi_z^-\rho\big)^{f_z^-} = \prod_{i,s} \left(\operatorname{tr}\Pi_i^s\rho\right)^{f_i^s} \ . \tag{S.6}$$

Taking the logarithm conveniently turns all the products into sums,

$$\ell\,(\rho) = \ln\mathcal{L}\,(\rho) = \sum_{i,s} f_i^s\, \ln\operatorname{tr}\left(\Pi_i^s\rho\right) \ . \tag{S.7}$$

It is now easy to verify that $\ell\,(\rho)$ is concave, using the concavity of the logarithm. Suppose that $\rho = \alpha\rho_1 + (1-\alpha)\rho_2$ for $0 \leqslant \alpha \leqslant 1$. Then

$$\begin{aligned}\ell\,(\rho) = \sum_{i,s} f_i^s\, \ln\operatorname{tr}\left(\Pi_i^s\,(\alpha\rho_1 + (1-\alpha)\rho_2)\right) &= \sum_{i,s} f_i^s\, \ln\left[\alpha\operatorname{tr}\left(\Pi_i^s\rho_1\right) + (1-\alpha)\operatorname{tr}\left(\Pi_i^s\rho_2\right)\right] \\ &\geqslant \sum_{i,s} f_i^s\, \left[\alpha\ln\operatorname{tr}\left(\Pi_i^s\rho_1\right) + (1-\alpha)\ln\operatorname{tr}\left(\Pi_i^s\rho_2\right)\right] = \alpha\,\ell\,(\rho_1) + (1-\alpha)\,\ell\,(\rho_2) \ .\end{aligned} \tag{S.8}$$

The inequality is a consequence of the concavity of the logarithm function.

(b) The *linear inversion* state ρ_{LI} is the one that reproduces the correct probabilities as given by the observed frequences. Calculate this state, which must satisfy

$$\operatorname{tr}\left(\rho_{\mathrm{LI}}\Pi_i^s\right) = \frac{f_i^s}{n_i} \ , \tag{2}$$

where Π_i^s is the POVM effect corresponding to measuring s on the observable σ_i (i.e., it is the projector onto the s eigenspace of σ_i). Is this state always well defined?

Hint. Go to the Bloch sphere picture.

Solution. We know from series 1 that in the Bloch sphere representation, measurements correspond to projecting your state onto one of two antipodal points of the sphere, with probabilities proportional to the projection of the point onto that measurement axis, as depicted in figure 2 (Left).

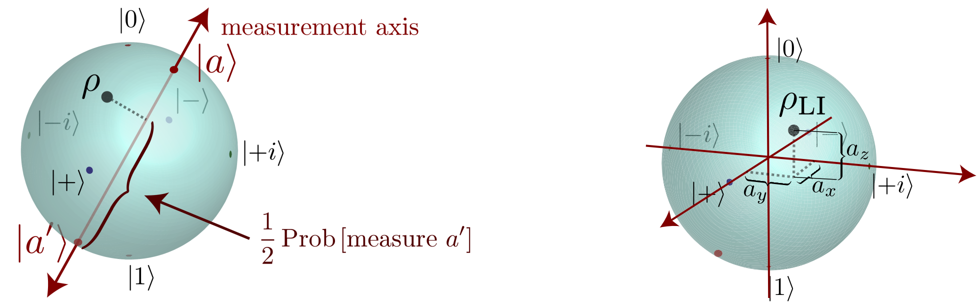


Figure 2: Tomography on the Bloch sphere. Left: we recall the representation of a quantum measurement in the Bloch sphere. Right: when we do linear inversion, we are looking for the Bloch sphere coordinates a_i^{LI} of ρ_{LI}, that match the required measurement probabilities for the corresponding measurements along the basis axes.

The measurement axes corresponding to the observables σ_x, σ_y and σ_z are simply the basis axes X, Y and Z on the Bloch sphere. So we are looking for a state ρ_{LI} with coordinates $\vec{a}^{\mathrm{LI}}$ on the Bloch sphere, that satisfies $\Pr\left[\text{“+” on } \sigma_i\right] = \frac{1}{2}\left(1 + a_i^{\mathrm{LI}}\right) = f_i^+/n_i$ for all $i = x, y, z$; see Fig. 2 (Right). So the vector $\vec{a}^{\mathrm{LI}}$ is simply given by

$$a_i^{\mathrm{LI}} = 2 \times \frac{f_i^+}{f_i^+ + f_i^-} - 1 = \frac{f_i^+ - f_i^-}{f_i^+ + f_i^-} \; . \tag{S.9}$$

(Note: the information about probabilities of measuring “−” are here redundant since for each σ_i we have $\Pr\left[\text{“}+\text{”}\right] = 1 - \Pr\left[\text{“}-\text{”}\right]$; consequently f_i^- appears only if we expand n_i as we did above as $f_i^+ + f_i^-$.)

Note that for the vector to lie inside the Bloch sphere, we must have $|\vec{a}^{\mathrm{LI}}| \leqslant 1$. However, since measurements are a probabilistic process and we only perform a finite number of measurements, it may be that $(f_i^+)/(f_i^+ + f_i^-)$ is not a good approximation of the true probability to measure “+” on observable σ_i, and since these errors are independent on all three axes, it may be in the end that $\vec{a}^{\mathrm{LI}}$ lies actually *outside* the Bloch sphere.

As a simple example, consider a fully mixed qubit (in the state $\mathbb{1}_2/2$) that you measure once only in each direction X, Y and Z. Suppose you get three times the outcome “+” (without loss of generality: the following conclusions stay the same for all the other measurement outcomes, too). Then, your linear inversion estimate would report $a_i = 1 \; \forall i$, which is clearly outside of the Bloch sphere.

This is actually one of the main problems of linear inversion estimation, which is dealt with by considering for example other estimation procedures like maximum likelihood.

(c) Let ρ_{MLE} be the maximum likelihood estimate. Show that if the linear inversion gives a state ρ_{LI} inside the state space, it coincides with the maximum likelihood estimate ρ_{MLE}.

Solution. Let us calculate the maximum likelihood estimate state vector $\vec{a}^{\mathrm{MLE}}$. Assume first that the maximum lies in the interior of the Bloch sphere. Then we can simply cancel the partial derivatives of $\ell(\vec{a})$,[1]

$$\frac{\partial}{\partial a_x}\ell(\vec{a}) = 0 \; ; \qquad \frac{\partial}{\partial a_y}\ell(\vec{a}) = 0 \; ; \qquad \frac{\partial}{\partial a_z}\ell(\vec{a}) = 0 \; . \tag{S.10}$$

Expressions (S.7) in the Bloch representation is

$$\ell(\vec{a}) = \sum_i \left[f_i^+ \ln\left(\frac{1 + a_i}{2}\right) + f_i^- \ln\left(\frac{1 - a_i}{2}\right)\right] \; , \tag{S.11}$$

recalling that $\operatorname{tr}\Pi_i^+\rho = (1 + a_i)/2$ and $\operatorname{tr}\Pi_i^-\rho = 1 - \operatorname{tr}\Pi_i^+\rho = (1 - a_i)/2$ are the probabilities, respectively, of observing “+” or “−” when measuring σ_i on state ρ. Then

$$\frac{\partial}{\partial a_i}\ell(\vec{a}) = f_i^+ \cdot \frac{1}{1 + a_i} - f_i^- \cdot \frac{1}{1 - a_i} \; . \tag{S.12}$$

Requiring that $(\partial\ell)/(\partial a_i) = 0$ then gives us

$$0 = f_i^+ (1 - a_i) - f_i^- (1 + a_i) = \left(f_i^+ - f_i^-\right) - a_i \left(f_i^+ + f_i^-\right) \; , \tag{S.13}$$

from which we deduce, in agreement with (S.9),

$$a_i = \frac{f_i^+ - f_i^-}{f_i^+ + f_i^-} \; . \tag{S.14}$$

[1] The solution, if any, is guaranteed to be the (unique) maximum because $\ell(\vec{a})$ is concave.

If, however, the maximum likelihood is on the border, then in general the linear inversion state can be outside the bloch sphere. One can understand handwavingly that the maximum likelihood estimate is the physical state that is "nearest" to the linear inversion state.

(In the case where the linear inversion state is also on the border, it still coincides with the maximum likelihood state, as one can see by running the above calculation backwards and showing that with $a_i = (f_i^+ - f_i^-)/(f_i^+ + f_i^-)$ the partial derivatives of $\ell(\vec{a})$ vanish.)

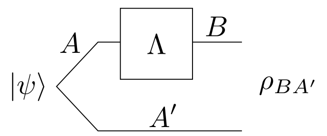
Exercise 3. ***Quantum process tomography***

Imagine that you want to characterize a quantum process (i.e., a TPCPM) $\Lambda_{A\to B}$. For instance, you were given a device by an untrusted party, and you want to verify that it actually does what you were told. The device takes as input a quantum state on system A and outputs states in system B. We will see that we can reduce process tomography to quantum state tomography.

Prepare a maximally entangled state between A and an ancilla system A',

$$|\psi\rangle = \frac{1}{|A|} \sum_i |i\rangle_A |i\rangle_{A'}, \tag{3}$$

and feed the part in A to your device, as displayed schematically in the following figure.



The resulting state is

$$\rho_{BA'} = [\Lambda_{A\to B} \otimes \mathcal{I}_{A'}](|\psi\rangle\langle\psi|_{AA'}). \tag{4}$$

Remark. Equation (4)*, seen as a mapping* $\Lambda_{A\to B} \mapsto \rho_{A'B}$*, can be shown to be an isomorphism mapping the completely positive, trace preserving maps to the density operators. This is known as the* Choi-Jamiolkowski isomorphism*.*

(a) Show that

$$\operatorname{tr}_{A'}[\rho_{BA'}(\mathbb{1}_B \otimes |k\rangle\langle\ell|_{A'})] = \frac{1}{|A|}\,\Lambda(|\ell\rangle\langle k|). \tag{5}$$

Solution. Using equation (4),

$$\begin{aligned}
\operatorname{tr}_{A'}[\rho_{BA'}(\mathbb{1}_B \otimes |k\rangle\langle\ell|_{A'})] &= \operatorname{tr}_{A'}\left[(\Lambda_{A\to B} \otimes \mathcal{I}_{A'})\,(|\psi\rangle\langle\psi|_{AA'}) \cdot (\mathbb{1}_B \otimes |k\rangle\langle\ell|_{A'})\right] \\
&= \operatorname{tr}_{A'}\left[\frac{1}{|A|}\sum_{i,j} (\Lambda_{A\to B} \otimes \mathcal{I}_{A'})\,(|i\rangle\langle j|_A \otimes |i\rangle\langle j|_{A'}) \cdot (\mathbb{1}_B \otimes |k\rangle\langle\ell|_{A'})\right] \\
&= \operatorname{tr}_{A'}\left[\frac{1}{|A|}\sum_{i,j} (\Lambda_{A\to B}\,(|i\rangle\langle j|) \otimes |i\rangle\langle j|_{A'}) \cdot (\mathbb{1}_B \otimes |k\rangle\langle\ell|_{A'})\right] \\
&= \frac{1}{|A|}\sum_{i,j} \Lambda_{A\to B}\,(|i\rangle\langle j|) \;\cdot \delta_{jk}\delta_{i\ell} \qquad = \frac{1}{|A|}\Lambda_{A\to B}\,(|\ell\rangle\langle k|) \;.
\end{aligned}$$

(b) Explain how you would proceed in order to obtain a full characterization of Λ, assuming that you could perform tomography on $\rho_{BA'}$.

Solution. Equation (5) tells us that the channel Λ is completely determined by $\rho_{BA'}$: if one is given $\rho_{BA'}$, then the output of Λ on an arbitrary input state σ_A can be calculated using linearity, matrix element by matrix element:

$$\begin{aligned}\Lambda(\sigma_A) &= \Lambda\Big(\sum_{k\,\ell} \langle k|\sigma_A|\ell\rangle\, |k\rangle\langle\ell|\Big) \\ &= \sum_{k\,\ell} \langle k|\sigma_A|\ell\rangle\, \Lambda\big(|k\rangle\langle\ell|\big) \\ &= |A|\, \sum_{k\,\ell} \langle k|\sigma_A|\ell\rangle\, \operatorname{tr}_{A'}\left[\rho_{BA'}\cdot(\mathbb{1}_B\otimes|\ell\rangle\langle k|_{A'})\right] \\ &= |A|\, \operatorname{tr}_{A'}\Big[\rho_{BA'}\cdot\Big(\mathbb{1}_B\otimes\sum_{k\,\ell}\langle k|\sigma_A|\ell\rangle\, |\ell\rangle\langle k|_{A'}\Big)\Big] \\ &= |A|\, \operatorname{tr}_{A'}\Big[\rho_{BA'}\cdot\Big(\mathbb{1}_B\otimes\sigma_{A'}^T\Big)\Big]\,,\end{aligned}$$

where $\sigma_{A'}^T$ is the transpose matrix of σ_A.

Thus if we're able to completely determine $\rho_{BA'}$, we can fully characterize Λ. However, we can simply perform the experiment explained in the exercise text, feeding one part of a maximally entangled state into Λ, and repeat this process many times in order to do full quantum state tomography jointly on the two output systems. This will determine $\rho_{BA'}$ completely (to arbitrary precision as the number of measurements increases), thus also determining the process Λ.