PHYS 610: Recent Developments in Quantum Mechanics and Quantum Information (Spring 2009)

Notes: Bayesian Statistics

1 A Third Prelude: Bayesian View of Quantum Measurement

With the introduction of POVMs as generalized measurements, we will now compare quantum measurements with classical Bayesian inference—gaining some insight into quantum measurements as processes of refining quantum information. We will only do so at a fairly simple level; more details on this modern view of quantum measurement may be gleaned from more extreme Bayesians.[1]

1.1 Bayes' Rule

Bayes' rule is the centerpiece of statistics from the Bayesian point of view, and is simple to derive. Starting from the definition of conditional probability $P(B|A)$, the probability of event B occuring *given* that event A occured,

$$P(A \wedge B) =: P(B|A)P(A), \tag{1}$$

we can use this to write what seems to be a simple formula:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}. \tag{2}$$

This is **Bayes' Rule**, which we will rewrite by replacing B by one possible outcome D_α out of a set $\{D_\alpha\}$ of all possible, disjoint outcomes:

$$P(A|D_\alpha) = \frac{P(D_\alpha|A)P(A)}{P(D_\alpha)}. \tag{3}$$

(Bayes' Rule)

Again, while this rule seems to be a fairly simple generalization of the definition of conditional probability, the key is in the *interpretation* of the various elements in this formula. The basic idea is that in learning that an outcome D_α actually occurred out of a set of possible measurement outcomes, $\{D_\alpha\}$ allows us to *refine* the probability we assign to A based on this new knowledge. The various factors are:

1. The **prior**: $P(A)$ represents the probability *assigned* to event A—*prior* to knowing the outcome of the measurement—based on any knowledge or assumptions. This probability is *not* conditioned on D_α.

2. The **probability of the measurement outcome**: $P(D_\alpha|A)$ is the probability that the particular measurement outcome, or event D_α would occur, given that A actually happened.

3. The **renormalization factor**: $P(D_\alpha)$ is the probability of the measurement outcome D_α, by which we must divide for the result to come out correctly. This is computed most simply by summing over the probabilities of a complete, nonintersecting set of outcomes A_β conditioned on D_α, weighted by the probabilities that the A_β occur:

$$P(D_\alpha) = \sum_\beta P(D_\alpha|A_\beta)P(A_\beta). \tag{4}$$

[1] Christopher A. Fuchs, "Quantum Mechanics as Quantum Information (and only a little more)," arXiv.org preprint quant-ph/0205039. This paper is interesting overall, but also worth reading *just* for the quote from Hideo Mabuchi on p. 13.

4. The **posterior**: $P(A|D_\alpha)$ is the *refined* probability of A, now that we know that the measurement outcome D_α has occurred.

The posterior probability thus reflects the *information gained* or revealed by the outcome event D_α.

1.2 Example: The "Monty Hall Problem"

One standard example of applying Bayes' rule is the **Monty Hall problem**. This is standard almost to the point of being painfully trite, but still this is a useful example in setting up our comparison to quantum measurement. We will define the rules as follows:

1. You're a contestant on the game show *Let's Make a Deal*, and you are shown three doors; we will call them doors 1, 2, and 3.
2. Behind one door is a brand-new car, and behind the other two are goats ("zonk prizes"). We will suppose that you like cars very much, but you aren't especially fond of goats: they smell funny and make you sneeze. We will also suppose that they are randomly placed, one behind each of the three doors, and the problem is invariant under any permutation of the door labels.
3. You pick a door; we will call that one "door 1" without loss of generality. You stand to gain whatever is behind that door.
4. The host opens up one of the other two doors to reveal a goat; without loss of generality we will call this door 3. We will assume the host knowingly and intentionally revealed a goat, and if he could do this in multiple ways, he would pick a door at random.
5. The problem is: is it to your advantage to switch to door 2, or should you stay with door 1?

The answer, somewhat counterintuitively, is that you double your chances of successfully winning the car if you switch doors. This result is not hard to work out using Bayes' rule:

- **Prior:** we will define the three events C_α, which is the event where the car is behind door α. Since the arrangement is random,
$$P(C_\alpha) = \frac{1}{3} \qquad (\forall_{\alpha\in\{1,2,3\}}). \tag{5}$$
- **Data:** the outcome event, or data, that gives us information is D_3, which will be our shorthand for the goat being behind door 3 *and* the host chose to reveal door 3 if there were multiple choices for revealing a goat. If the car were behind door 1, then there are goats behind doors 2 and 3, so the host would have a 50% chance of opening door 3:
$$P(D_3|C_1) = \frac{1}{2}. \tag{6}$$
If the car were behind door 2, then opening door 3 would be the only choice,
$$P(D_3|C_2) = 1, \tag{7}$$
while if the car were behind door 3, opening door 3 wouldn't be an option:
$$P(D_3|C_3) = 0. \tag{8}$$

The probability for D_3 to occur is given by summing over all conditional probabilities for D_3, weighted by the probability of each conditioning event to occur:

$$P(D_3) = \sum_\alpha P(D_3|C_\alpha)P(C_\alpha) = \frac{1}{2}\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} = \frac{1}{2}. \tag{9}$$

- **Posterior:** Now, given the information revealed by the host's choice, we can compute the posterior probabilities of the car being behind each door:

$$\begin{aligned} P(C_1|D_3) &= \frac{P(D_3|C_1)P(C_1)}{P(D_3)} = \frac{1}{3} \\ P(C_2|D_3) &= \frac{P(D_3|C_3)P(C_3)}{P(D_3)} = \frac{2}{3} \\ P(C_3|D_3) &= \frac{P(D_3|C_3)P(C_3)}{P(D_3)} = 0. \end{aligned} \tag{10}$$

Clearly it is to your advantage to switch to door 2, since the probability of finding the car there is double what it was before. Note that, in accounting for the action of the host, the probability distribution for finding the car behind each door changed discontinuously: the distribution was initially uniform, then changed to a different situation where one possibility has the maximum probability, and another possibility has been ruled out. This is quite reminiscent of wave-function collapse after a quantum measurement.

1.2.1 Quantum Language

In fact, we can recast this same problem in the notation of quantum-mechanical measurement as represented by POVMs quite easily. This is the identical problem, though, so there will in fact be nothing quantum-mechanical about the treatment of this example except for the notation. This is simply an exercise to emphasize the similarity of quantum measurements to Bayes' rule.

We will label the three outcomes by the states $|\alpha\rangle$, with projection operators $P_\alpha := |\alpha\rangle\langle\alpha|$. The initial state is equiprobable, and for a classical mixture the density operator is thus simply proportional to the identity operator:

$$\rho = \frac{1}{3}(P_1 + P_2 + P_3). \tag{11}$$

We can then represent the revelation of a goat behind door 3 by guessing the operator

$$\Omega := \frac{1}{\sqrt{2}}P_1 + P_2. \tag{12}$$

Thus, $\Omega^\dagger\Omega = \frac{1}{2}P_1 + P_2$, and a POVM could be completed, for example, by the alternate possibility $\Omega = \frac{1}{\sqrt{2}}P_1 + P_3$. We can verify that the operator Ω gives the right conditional probabilities for the each of the outcomes, given by the appropriate trace $\mathrm{Tr}[\Omega\rho\Omega^\dagger]$, setting the density operator equal to the appropriate projector, $\rho = P_\alpha$:

$$\begin{aligned} P(D_3|C_1) &= \mathrm{Tr}[\Omega P_1\Omega^\dagger] = \frac{1}{3}\mathrm{Tr}\left[\frac{P_1^2}{2}\right] = \frac{1}{2} \\ P(D_3|C_2) &= \mathrm{Tr}[\Omega P_2\Omega^\dagger] = \mathrm{Tr}[P_2^2] = 1 \\ P(D_3|C_3) &= \mathrm{Tr}[\Omega P_3\Omega^\dagger] = 0. \end{aligned} \tag{13}$$

These are, of course, the same classical probabilities as in Eqs. (6)–(8), and this is precisely the justification for defining this operator. Now the **conditioned state** ρ_{c} is given by the POVM transformation

$$\rho_{\mathrm{c}} = \frac{\Omega\rho\Omega^\dagger}{\mathrm{Tr}[\Omega\rho\Omega^\dagger]} = \frac{\left(\frac{P_1}{\sqrt{2}} + P_2\right)\frac{1}{3}(P_1+P_2+P_3)\left(\frac{P_1}{\sqrt{2}} + P_2\right)}{\mathrm{Tr}\left[\left(\frac{P_1}{\sqrt{2}} + P_2\right)\frac{1}{3}(P_1+P_2+P_3)\left(\frac{P_1}{\sqrt{2}} + P_2\right)\right]} = \frac{\left(\frac{P_1}{2} + P_2\right)}{\mathrm{Tr}\left[\frac{P_1}{2} + P_2\right]} = \frac{1}{3}P_1 + \frac{2}{3}P_2. \tag{14}$$

Finally, the posterior, or conditioned, probabilities of finding the car behind each of the doors is given by a similar trace, where the projector P_α defines the outcome of finding the car behind door α in a *future* measurement:

$$\begin{aligned} P_{\mathrm{c}}(C_1) &= \mathrm{Tr}[P_1\rho_{\mathrm{c}}P_1] = \frac{1}{3} \\ P_{\mathrm{c}}(C_2) &= \mathrm{Tr}[P_2\rho_{\mathrm{c}}P_2] = \frac{2}{3} \\ P_{\mathrm{c}}(C_3) &= \mathrm{Tr}[P_3\rho_{\mathrm{c}}P_3] = 0. \end{aligned} \tag{15}$$

These are the same probabilities that we obtained using Bayes' rule in standard form.

1.3 Quantum Measurement as Inference from Data

To generalize the Monty Hall example, we can recast the POVM reduction as a "quantum Bayes' Rule." Assume we have a set D_α of Krause operators that are comprised in a POVM. Then the αth measurement outcome converts the quantum state ρ into the conditioned state ρ_{c} according to

$$\rho_{\mathrm{c}} = \frac{D_\alpha\rho D_\alpha^\dagger}{\mathrm{Tr}[D_\alpha\rho D_\alpha^\dagger]}. \tag{16}$$

We can identify elements here that are very similar to the classical Bayes' Rule:

1. The **prior**: in this case is the initial density operator ρ.
2. The **reduction**: the operators D_α and $D_\alpha^\dagger$ act like the conditional probability $P(D_\alpha|A)$ in the classical case, which effects the change in the probability in response to the occurrence of D_α (regarded as an event). As we saw in the Monty Hall example, these quantum operators can be constructed to be equivalent to the classical conditional probabilities.
3. The **renormalization factor**: we then renormalize the probability by dividing by $\mathrm{Tr}[D_\alpha\rho D_\alpha^\dagger]$, which is just the probability $P(D_\alpha)$ in the classical case. This step of course ensures a normalized, conditioned density operator, which we of course need for a sensible probability distribution.
4. The **posterior**: the conditioned state ρ_{c} then reflects our knowledge of the quantum state given the αth outcome of the measurement, in the same way that $P(A|D_\alpha)$ reflects the probability for outcome A given the event D_α.

The obvious but superficial difference here is that the classical rule describes the change in the assigned probability for a *single* event A, whereas the quantum rule handles *all possible* outcomes of a future measurement all at once. While similar, the quantum and classical rules can't quite be cast in the same form since the quantum rule is both more general in handling *coherent* superpositions (quantum probabilities) and different in that measurements on some aspects of a system must disturb complementary aspects (quantum backaction). We can conclude this interlude by noting a number of points regarding how one can use the quantum Bayes' rule as a framework for thinking about quantum measurement.[2] While bordering on the philosophical, this is a very useful framework for thinking about measurements in modern experiments, particularly where single quantum systems and multiple, sequential measurements are involved.

- The quantum state ρ *is* the information about a quantum system according to a particular observer.
- A quantum measurement *refines* the observer's information about the system, and thus modifies the density operator.
- This removes any problems with "collapse of the wave function" as a discontinuous process. The wave function is, in fact, literally in the observer's head, and the collapse is just an update of information.
- This view is particularly useful in considering multiple observers for the same system, both performing their own weak measurements but possibly not sharing their results. We treated this, for example, in the case of stochastic master equations for photodetection, where each observer ends up with a different *conditioned* density operator. Each density operator incorporates the measurement information of the corresponding observer, but also a trace over the unknown measurement results of the *other* observer.
- The price of all this is the potentially distasteful feature of *subjective*, or *observer-dependent* quantum states. Actually, this shouldn't be unreasonable in the case of multiple observers; however, even multiple observers with access to *all* the same measurement results could disagree on the details of the quantum state, because they may have begun with different prior states. There *are* a number of important objective features, however. For example, the data (measurement results) are objective—as are the rules for incorporating data—and as the observers continue to incorporate more data, their states should converge (at least in the aspects reflected by the measurements): with sufficient data, eventually the information from the data should completely swamp the prior. Further, in constructing priors, both observers should either agree that the probability of a particular measurement outcome is either zero or nonzero, even if they disagree on the exact probability: an assigned probability of zero is the only really claim that is absolutely falsifiable by future measurements. Finally, there are objective ways of constructing prior states, such as the *maximum-entropy principle*, which chooses the state with the least information that is consistent with all known constraints[3] (though in practice determining and implementing constraints can be a tricky business[4]).
- In any quantum or classical measurement, the knowledge should increase, or at least it shouldn't *decrease*. For example, in a quantum projective measurement, a mixed state always transforms to a pure state, with correspondingly less uncertainty (i.e., the uncertainty reduced to the quantum-mechanical minimum), though of course once in a pure state a projection can only modify the state without increasing knowledge. Essentially the same is true of general POVMs.[5]
- The information gained in a quantum measurement is *not* about some pre-existing reality (i.e., hidden variables), but rather in the measurement, the uncertainty *for predictions of future measurements* decreases.

2 Exercise

Suppose that a (fictitious) serious disease called Bayes' syndrome affects 0.1% of the population. Suppose also that you are tested for Bayes' syndrome. The test has a false-positive rate of 0.1% and a false-negative

[2]see Christopher A. Fuchs, *op. cit.* for *much* additional detail.

[3]E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge, 2003).

[4]Jos Uffink, "The Constraint Rule of the Maximum Entropy Principle," *Studies in History and Philosophy of Modern Physics* **27**, 47 (1996) (doi: 10.1016/1355-2198(95)00022-4).

[5]see Christopher A. Fuchs, *op. cit.* for proofs.

rate of 0.1% (i.e., for either outcome the test is correct 99.9% of the time). If you test positive for Bayes' syndrome, what is the probability that you actually have the disease?

Use Bayes' rule to calculate the answer, identifying the various factors in Bayes' rule. Surprisingly, the answer turns out to be 50%. The reason is that the prior knowledge of the situation is that you are very unlikely to have the disease, and because it is so skewed, the prior strongly influences the posterior expectation.