Toward a fully quantum version of Bayes' rule

Ge Bai, Francesco Buscemi*, Valerio Scarani

*Department of Mathematical Informatics, Nagoya University

Mathematics of Quantum Information, Aachen, Germany

Abstract

Although Bayes' rule is often considered a trivial identity, its effectiveness as an updating rule warrants further investigation. In this talk, I revisit the classical Bayesian framework through the lens of the minimum change principle. This principle asks that we update our beliefs to incorporate new data while deviating minimally from the prior. I demonstrate how this principle naturally leads to the generalizations of Bayes' rule proposed by Jeffrey and Pearl in noisy classical scenarios. I then extend this idea to quantum systems and present a fully quantum update rule that recovers the Petz transpose map as the unique solution to a quantum optimization task. This establishes the Petz map as more than just a mathematical artifact; it is also a principled quantum analog of Bayesian inference.

Toward a fully quantum version of Bayes' rule

Francesco Buscemi, Nagoya University

Mathematics of Quantum Information
11 July 2025, Aachen, Germany

works this talk is based upon

- F. Buscemi and V. Scarani, *Fluctuation theorems from Bayesian retrodiction*. Physical Review E, vol. 103, 052111 (2021).
- C.C. Aw, F. Buscemi, and V. Scarani, *Fluctuation theorems with retrodiction rather than reverse processes*. AVS Quantum Science, vol. 3, 045601 (2021).
- F. Buscemi, J. Schindler, and D. Šafránek, *Observational entropy, coarse-grained states, and the Petz recovery map: information-theoretic properties and bounds*. New Journal of Physics, vol. 25, 053002 (2023).
- G. Bai, D. Šafránek, J. Schindler, F. Buscemi, and V. Scarani, *Observational entropy with general quantum priors*. Quantum, vol. 8, 1524 (2024).
- G. Bai, F. Buscemi, and V. Scarani, *Quantum Bayes' rule and Petz transpose map from the minimum change principle*. Preprint arXiv:2410.00319 (2024).
- G. Bai, F. Buscemi, and V. Scarani, *Fully quantum stochastic entropy production*. Preprint arXiv:2412.12489 (2024).
- T. Nagasawa, K. Kato, E. Wakakuwa, and F. Buscemi, *Macroscopic states and operations: a generalized resource theory of coherence*. Preprint arXiv:2504.12738 (2025).

what this talk is not about

philosophical debates (e.g., Bayesianism VS Frequentism), interpretations of QM (e.g., QBism), etc.

we are postmodern Bayesians!

what this talk is about

$$\underbrace{P(H|D)}_{\text{posterior}} = \frac{\overbrace{P(H)}^{\text{prior}}\overbrace{P(D|H)}^{\text{likelihood}}}{\underbrace{P(D)}_{\text{prop. constant}}}$$

This talk is about Bayes' rule and its "unreasonable effectiveness"

The **wrong answer**: it is a trivial consequence of the law of total probability, detailed balance, etc.

The **correct question**: why Bayes' rule provides a good update rule?

possible justifications of the effectiveness of Bayes' rule

"Consistency" arguments by De Finetti, (Harold) Jeffreys, Savage, and Cox.

(Richard) Jeffrey's "probability kinematics" and Pearl's "virtual evidence method".

parenthesis: the Bayes–Jeffrey–Pearl update

- consider a classical discrete noisy channel $P(i|x)$ and a prior $\gamma(x)$ on the input
- when the receiver reads a definite value i_0, (vanilla) Bayes' rule says that their posterior should be updated to $R_P^\gamma(x|i_0) := \dfrac{\gamma(x)P(i_0|x)}{[P\gamma](i_0)}$
- but what if the observation is noisy and returns some p.d. $\sigma(i)$ instead?

Theorem (Jeffrey 1965, Pearl 1988)

Given a channel $P(i|x)$ and a prior $\gamma(x)$, the result of a noisy observation $\sigma(i)$ is updated to

$$\widetilde{\sigma}(x) := \sum_i \boxed{R_P^\gamma(x|i)}\, \sigma(i)\ .$$

Note: the usual Bayes' rule is recovered for $\sigma(i) = \delta_{i,i_0}$.

The problem with these derivations is that they are based on axioms, which may appear compelling to some but less so to others.

Alternative approach: can Bayes' rule be derived as the (optimal, unique) solution to a concrete task?

the conservative stance

To avoid unwarranted bias and remain maximally non-committal, the updated belief should be consistent with the new information (the result of the observation), while deviating as little as possible from the initial belief.

formalization: the principle of minimum change

How Bayes' rule can be derived from this principle:

- given channel $P(i|x)$ and prior $\gamma(x)$, construct the forward process $[P\star\gamma](i,x) := P(i|x)\gamma(x)$
- given the new data as $\sigma(i)$, consider the program

$$\min_R \mathbb{D}(P\star\gamma, R\star\sigma)\ ,$$

  where $\mathbb{D}(\bullet,\bullet)$ is a suitable information divergence, and the minimum is taken over all channels $R\equiv R(x|i)$

Then, for many reasonable choices of $\mathbb{D}$ (e.g., the KL-divergence), it turns out that

$$\arg\min_R \mathbb{D}(P\star\gamma, R\star\sigma) = \{R_P^\gamma\}\ ,$$

where $R_P^\gamma \equiv R_P^\gamma(x|i) = \frac{[P\star\gamma](i,x)}{[P\gamma](i)}$ is Bayes' inverse.

towards a quantum generalization

The minimum change principle is formulated using the *joint* input-output distributions.

Hence, the central idea is that the "change" to be minimized is the change relative to the *whole input-output stochastic process*, not just its marginals.

But this is a **problem** in the quantum case...

quantum joint distributions

Given a channel $\mathcal{E} : A \to B$ define:

- the Choi operator: $C_{\mathcal{E}} := \sum_{i,j} \mathcal{E}(|i\rangle\langle j|)_B \otimes |i\rangle\langle j|_A$
- the joint state: $\mathcal{E} \star \gamma := (\mathbb{1}_B \otimes \sqrt{\gamma_A^T}) C_{\mathcal{E}} (\mathbb{1}_B \otimes \sqrt{\gamma_A^T})$

Note that:

- $\mathrm{Tr}_B[\mathcal{E} \star \gamma] = \gamma_A^T$
- $\mathrm{Tr}_A[\mathcal{E} \star \gamma] = \mathcal{E}(\gamma)_B$
- when all operators are diagonal, we obtain the classical input-output joint probability distribution

the principle of minimum change: the quantum case

Given a channel $\mathcal{E} : A \to B$, a prior state γ_A some "new information" σ_B, consider the program

$$\min_{\mathcal{R}} \mathbb{D}(\mathcal{E} \star \gamma, (\mathcal{R} \star \sigma)^T) ,$$

where the minimum is taken over channels $\mathcal{R} : B \to A$.

Theorem (arXiv:2410.00319)

For $\mathcal{E} \star \gamma > 0$ and $\sigma > 0$, when the divergence is chosen to be the quantum fidelity,

$$\arg\min_{\mathcal{R}} \mathbb{D}(\mathcal{E} \star \gamma, (\mathcal{R} \star \sigma)^T) = \{\mathcal{P}_{\mathcal{E},\gamma,\sigma}\} ,$$

where

$$\mathcal{P}_{\mathcal{E},\gamma,\sigma}(\bullet) := \sqrt{\gamma}\, \mathcal{E}^\dagger \left(\sqrt{\sigma} \frac{1}{\sqrt{\sqrt{\sigma}\mathcal{E}(\gamma)\sqrt{\sigma}}} \bullet \frac{1}{\sqrt{\sqrt{\sigma}\mathcal{E}(\gamma)\sqrt{\sigma}}} \sqrt{\sigma} \right) \sqrt{\gamma} .$$

Note: $[\mathcal{E}(\gamma), \sigma] = 0 \implies \mathcal{P}_{\mathcal{E},\gamma,\sigma} \equiv \mathcal{P}_{\mathcal{E},\gamma}$ coincides with Petz's transpose map.

concluding remarks

In general, when $[\mathcal{E}(\gamma), \sigma] \neq 0$ the dependence of the posterior $\mathcal{P}_{\mathcal{E},\gamma,\sigma}(\sigma)$ on the new data σ is not linear: bug or feature?

Useful application: when $\mathcal{E}$ is a qc-channel $\mathcal{E}(\bullet) := \sum_x \mathrm{Tr}[\bullet\, E_x]\, |x\rangle\langle x|$, Petz's transpose map is the uniquely optimal retrodiction that can be drawn from the measurement outcomes.

In particular, the "mysterious" Gibbsian coarse-graining is nothing but retrodiction done upon observation: $\mathcal{C}_{\mathcal{E},\gamma}(\bullet) := [\mathcal{R}_{\mathcal{E},\gamma} \circ \mathcal{E}](\bullet) = \sum_x \mathrm{Tr}[\bullet\, E_x]\, \frac{\sqrt{\gamma} E_x \sqrt{\gamma}}{\mathrm{Tr}[\gamma E_x]}$.

What happens when other divergences are used instead of the fidelity?

What about multi-partite situations, locality restrictions, ...?

Happy birthday, Mario!

References

Main:

- G. Bai, F. Buscemi, and V. Scarani, *Quantum Bayes' rule and Petz transpose map from the minimal change principle.* Preprint arXiv:2410.00319 (2024).

Other references:

- F. Buscemi and V. Scarani, *Fluctuation theorems from Bayesian retrodiction.* Physical Review E, vol. 103, 052111 (2021).
- C.C. Aw, F. Buscemi, and V. Scarani, *Fluctuation theorems with retrodiction rather than reverse processes.* AVS Quantum Science, vol. 3, 045601 (2021).
- F. Buscemi, J. Schindler, and D. Šafránek, *Observational entropy, coarse-grained states, and the Petz recovery map: information-theoretic properties and bounds.* New Journal of Physics, vol. 25, 053002 (2023).
- G. Bai, D. Šafránek, J. Schindler, F. Buscemi, and V. Scarani, *Observational entropy with general quantum priors.* Quantum, vol. 8, 1524 (2024).
- T. Nagasawa, K. Kato, E. Wakakuwa, and F. Buscemi, *On the generic increase of observational entropy in isolated systems.* Physical Review Research, vol. 6, 043327 (2025).
- G. Bai, F. Buscemi, and V. Scarani, *Quantum Bayes' rule and Petz transpose map from the minimal change principle.* Preprint arXiv:2410.00319 (2024).
- G. Bai, F. Buscemi, and V. Scarani, *Fully quantum stochastic entropy production.* Preprint arXiv:2412.12489 (2024).
- T. Nagasawa, K. Kato, E. Wakakuwa, and F. Buscemi, *Macroscopic states and operations: a generalized resource theory of coherence.* Preprint arXiv:2504.12738 (2025).