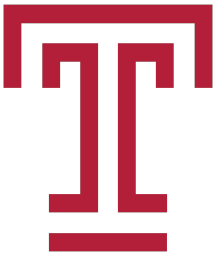
Artificial Intelligence

CIS 5603

Quantum-Accelerated Bayesian Inference for Rejection Sampling

Murathan Akman



15 December 2025

Contents

1 Introduction

Bayesian Networks (BNs) provide a principled framework for representing uncertainty and causal structure in complex systems. They are widely used in areas such as medical diagnosis, decision support systems, and probabilistic reasoning under uncertainty. A Bayesian Network encodes joint probability distributions over a set of random variables using a directed acyclic graph (DAG), where nodes represent variables and edges represent conditional dependencies.

Despite their expressive power, exact inference in Bayesian Networks becomes computationally intractable as the size of the network grows or when only partial information is available. In practical settings, one rarely queries the full joint distribution. Instead, the objective is often to compute conditional or marginal probabilities given some observed evidence. This requirement introduces the need for marginalization, which is the main source of computational complexity in Bayesian inference.

This report focuses on the fundamental limitations of classical inference methods in the presence of rare evidence and motivates the need for alternative approaches, which later sections will connect to quantum-based inference techniques.

2 Classical Bayesian Inference Foundations

2.1 Conditional Probability and Bayesian Networks

In probabilistic reasoning, we often care about how the probability of one event changes when we observe another event. This is captured by conditional probability:

$$P(A \mid B) = \frac{P(A, B)}{P(B)} \tag{1}$$

Bayesian Networks represent joint probability distributions in a structured way using a directed graph. The key benefit is that the full joint distribution can be factorized into local conditional probabilities, one per node conditioned on its parents:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i)) \tag{2}$$

This factorization is powerful because it replaces an unstructured joint table with a product of smaller conditional probability tables (CPTs).

2.2 Marginalization and Sum-Out Operations

In practice, we rarely need the full joint distribution over all variables. Instead, we want a marginal or conditional probability of only some variables, while other variables remain unknown or irrelevant to the query. In those cases, we must sum out (marginalize) the variables that we do not care about.

For example, if we want the distribution of X_1 given x_2, x_3, but x_4, x_5 are not observed or not part of the query, we compute:

$$P(X_1 \mid x_2, x_3) = \sum_{x_4} \sum_{x_5} P(x_1, x_2, x_3, x_4, x_5) \tag{3}$$

This is the core idea behind variable elimination or sum-out: unknown variables must be aggregated away to obtain the probability of interest.

2.3 Experimental Bayesian Network Scenario

This section describes the Bayesian Network used in the experimental setup, including its structure, prior distributions, and conditional probability tables.

Bayesian Network Structure

The Bayesian Network consists of three binary variables:

- L: latent variable representing a rare event
- A: auxiliary variable
- R: observed result variable

Figure 1 illustrates the structure of the Bayesian Network used in the experiments.

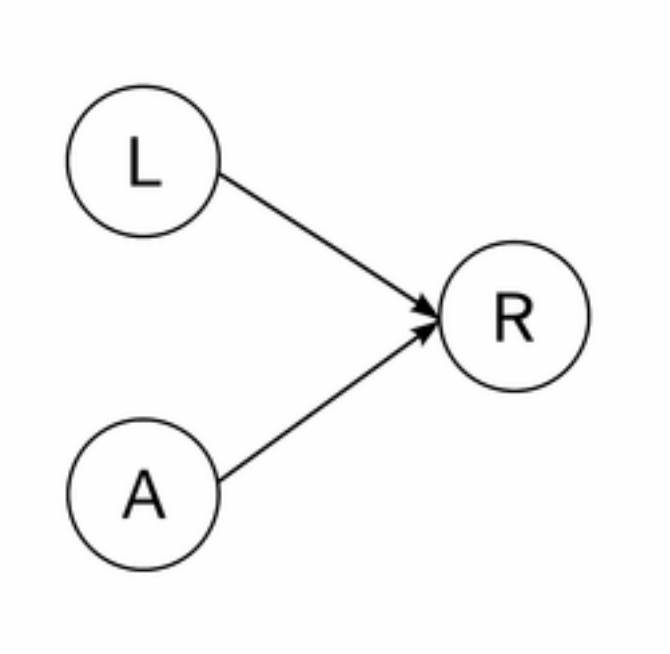


Figure 1: Bayesian Network structure for the experimental scenario

Prior Distributions

The prior probability distributions are defined as follows:

- Prior of latent variable L:
  - $P(L = 1) = 0.001$
  - $P(L = 0) = 0.999$
- Prior of auxiliary variable A:
  - $P(A = 1) = 0.5$
  - $P(A = 0) = 0.5$

Conditional Probability Table

The conditional probability table (CPT) for the variable R given its parents L and A is shown in Table 1.

Table 1: Conditional Probability Table for $P(R \mid L, A)$

L	A	$P(R = 1 \mid L, A)$	$P(R = 0 \mid L, A)$
0	0	$3/5 = 0.6$	$2/5 = 0.4$
0	1	$1/5 = 0.2$	$4/5 = 0.8$
1	0	$1/8 = 0.125$	$7/8 = 0.875$
1	1	$4/7 \approx 0.571428$	$3/7 \approx 0.428572$

3 Approximate Classical Inference

3.1 Rejection Sampling Intuition

When exact inference becomes expensive, approximate inference methods are used. Rejection Sampling is one such method. The idea is:

- Generate full samples from the Bayesian Network using its CPTs
- Keep (accept) only the samples that match the evidence
- Discard the rest

A key point is that a particular configuration of variable values is sampled with its corresponding probability:

$$P(\text{sample} = \langle X_1 = \text{true}, X_2 = \text{false} \rangle) = P(X_1 = \text{true}, X_2 = \text{false}) \tag{4}$$

So, if an event is rare, the sampler will almost never generate itmeaning many samples get wasted.

3.2 Estimating Conditional Probabilities from Samples

Using samples, we can approximate conditional probabilities by counting:

$$P(X_1 = \text{true} \mid X_2 = \text{false}) \approx \frac{\#Samples(X_1 = \text{true}, X_2 = \text{false})}{\#Samples(X_2 = \text{false})} \tag{5}$$

This estimate improves as we collect more samples. However, the critical bottleneck appears when the evidence itself is rare.

3.3 The Rare Evidence Problem

If the evidence probability is extremely small (e.g., $P(L = 1) = 0.001$), then to obtain enough accepted samples (e.g., 100 accepted samples), we may need on the order of $\sim$ 100,000 total generated samples, because almost all samples are discarded. This is exactly the rare evidence problem: the method becomes inefficient because acceptance is too unlikely.

4 Bayesian Decision-Making

4.1 Expected Utility Framework

A Bayesian Network is not only for computing probabilities; it can also support decision making. For decision making, we include a utility function that represents how good or bad an outcome is. Each possible result has:

- a probability
- a utility (value or benefit)

Expected Utility is defined as probability-weighted utility summed over possible results:

$$EU(a \mid e) = \sum_r P(\text{Result} = r \mid a, e) \cdot U(r) \tag{6}$$

This means: to compute how beneficial an action a is (given evidence e):

- consider each possible outcome r
- multiply its probability $P(r \mid a, e)$ by its utility $U(r)$
- sum them all

A concrete example is deciding whether to take a blood test (test vs no test). If action $a =$ "take the test", possible results might be:

1. R_1: cancer found early
2. R_2: nothing found
3. R_3: false positive

Each has a probability $P(R_i \mid a, e)$ and a utility $U(R_i)$. Then:

$$EU(\text{test}) = P(R_1) \cdot U(R_1) + P(R_2) \cdot U(R_2) + P(R_3) \cdot U(R_3) \tag{7}$$

The important point is: probability alone does not decide. Even if something is low probability, it might have very high utility (e.g., early detection). Bayesian inference helps estimate what is likely; utility tells what is valuable; expected utility helps select the best decision.

5 Quantum Foundations for Bayesian Inference

5.1 Quantum States and Unitary Evolution

Quantum algorithms operate by manipulating quantum states through unitary transformations. A quantum state is represented as a complex-valued vector whose squared amplitudes correspond to measurement probabilities. For a single qubit, the state can be written as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{8}$$

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$.

This normalization condition ensures that the total probability of all possible measurement outcomes remains conserved. Quantum evolution is governed exclusively by unitary matrices, which preserve this norm. Every quantum gate applied to a qubit corresponds to a unitary matrix, and a quantum circuit is therefore the ordered product of unitary transformations applied to the initial state.
From a computational perspective, a quantum algorithm can be viewed as a carefully designed unitary operator that reshapes the amplitude distribution of the quantum state in a goal-oriented manner. Importantly, until measurement occurs, this evolution is fully reversible.

6 Circuit Representation

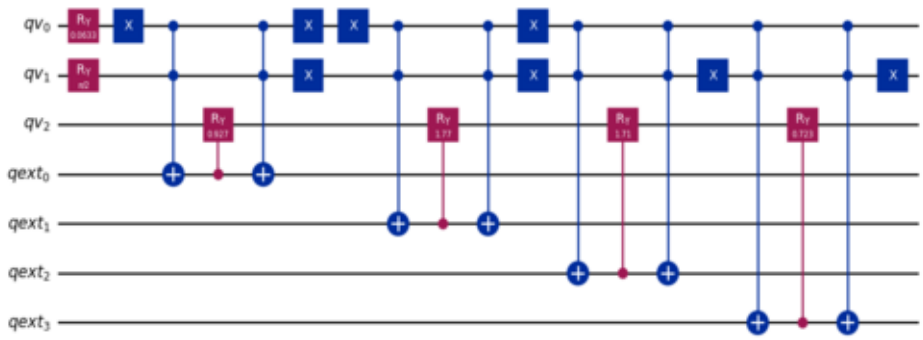


Figure 2: Circuit Design with Quantum Gates

This normalization condition ensures that the total probability of all possible measurement outcomes remains conserved. Quantum evolution is governed exclusively by unitary matrices, which preserve this norm. Every quantum gate applied to a qubit corresponds to a unitary matrix, and a quantum circuit is therefore the ordered product of unitary transformations applied to the initial state.
From a computational perspective, a quantum algorithm can be viewed as a carefully designed unitary operator that reshapes the amplitude distribution of the quantum state in a goal-oriented manner. Importantly, until measurement occurs, this evolution is fully reversible.

7 From Classical Rejection Sampling to Quantum Bayesian Rejection Sampling

Classical rejection sampling operates by repeatedly generating random samples from a joint probability distribution and discarding those that do not satisfy a given evidence condition. While conceptually simple, this approach becomes extremely inefficient when the evidence is rare. If an event has probability p, then approximately $O(1/p)$ samples are required to obtain a fixed number of valid samples.
The quantum version of rejection sampling leverages two uniquely quantum resources:

1. Superposition, which allows all possible assignments of variables to be represented simultaneously.

2. Amplitude amplification, which increases the probability of measuring desired outcomes while suppressing unwanted ones.

Instead of discarding samples, the quantum algorithm redistributes probability mass through interference. As a result, the number of oracle applications required to amplify rare events scales as $O(1/\sqrt{p})$, yielding a quadratic speedup over classical rejection sampling. This behavior is directly analogous to Grovers search algorithm.

8 Encoding Bayesian Networks into Quantum States

In the quantum Bayesian framework, each node of a Bayesian network is encoded into the amplitudes of a qubit. For a binary random variable X with probabilities

$$P(X = \text{true}) = p, \quad P(X = \text{false}) = 1 - p, \tag{9}$$

the corresponding quantum state is

$$|\psi_X\rangle = \sqrt{p}\,|0\rangle + \sqrt{1-p}\,|1\rangle. \tag{10}$$

This encoding ensures that measurement probabilities match the classical distribution exactly. More importantly, when multiple variables are combined, the resulting multi-qubit state naturally represents the joint probability distribution, with amplitudes encoding correlations specified by the Bayesian networks conditional probability tables (CPTs).
This approach allows the entire CPT structure of a Bayesian network to be embedded into a quantum state, enabling global operations on the full joint distribution rather than local sampling-based updates.

9 Density Matrix Representation and Quantum Coherence

While state vectors are sufficient for pure quantum states, a more general and informative representation is given by the density matrix

$$\rho = |\psi\rangle\langle\psi|. \tag{11}$$

For a single qubit with state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the density matrix takes the form

$$\rho = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}. \tag{12}$$

The diagonal elements correspond to classical probabilities, while the off-diagonal elements represent quantum coherence. These coherence terms capture phase relationships between states and are responsible for interference effectsphenomena with no classical analogue.
In the context of Bayesian inference, coherence enables the quantum algorithm to exploit relationships between probabilistic configurations. Destructive interference suppresses undesirable outcomes, while constructive interference amplifies desired ones. This mechanism is fundamental to the performance advantage of quantum rejection sampling.

10 Interference, Amplitude Amplification, and Quantum Advantage

Quantum interference plays a central role in eliminating unwanted probabilistic configurations. In classical inference, all configurations must be examined explicitly or filtered probabilistically. In contrast, quantum algorithms allow certain configurations to be cancelled out via destructive interference.
Two major advantages arise from this mechanism:

1. Suppression of unwanted outcomes Configurations inconsistent with the evidence experience amplitude cancellation, reducing their likelihood of being measured.

2. Amplification of desired outcomes Through amplitude amplification, valid samples are reinforced via constructive interference, significantly increasing their measurement probability.

This process underlies the quadratic speedup observed in quantum Bayesian rejection sampling. Importantly, the strength of interferenceand therefore the quantum advantagedepends on the coherence of the system. As coherence diminishes due to noise or decoherence, quantum behavior gradually approaches classical sampling.
These observations clarify why quantum advantage is fundamentally linked to coherence and why noisy hardware can limit achievable speedups in practice.

11 Concrete Quantum Example: Phase Encoding and Interference

To make the abstract discussion concrete, we consider a minimal quantum example involving two qubits, which can be interpreted as a classical Bayesian node and its child. In this simplified setting, the computational basis states are interpreted as follows:

- $|00\rangle$: "good" state
- $|11\rangle$: "bad" state

The corresponding quantum state is written as

$$|\Psi\rangle = \alpha|00\rangle + \beta|11\rangle \tag{13}$$

where α and β are real amplitudes such that $|\alpha|^2 + |\beta|^2 = 1$.
The coherence between these two components is given by

$$\rho_{00,11} = \alpha\beta^*. \tag{14}$$

For example, if $\alpha = \sqrt{0.6}$ and $\beta = \sqrt{0.4}$, then the coherence becomes

$$\rho_{00,11} = \sqrt{0.6} \times \sqrt{0.4} \approx 0.49. \tag{15}$$

To mark the undesired (bad) state, we apply a phase flip using the Pauli-Z gate:

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{16}$$

Applied conditionally, this transformation yields

$$|00\rangle \to |00\rangle, \quad |11\rangle \to -|11\rangle.$$

As a result, the quantum state becomes

$$|\Psi\rangle = \alpha|00\rangle - \beta|11\rangle. \tag{17}$$

This phase marking is the quantum analogue of identifying bad samples in classical rejection sampling.

12 Oracle Interpretation and the Role of Interference

The oracle is defined as a unitary operator that flips the phase of configurations that violate the evidence condition:

$$O|x\rangle = \begin{cases} -|x\rangle & \text{if } x \text{ is bad} \\ |x\rangle & \text{otherwise.} \end{cases} \tag{18}$$

After the oracle, interference redistributes amplitude mass: amplitudes associated with good states increase, while amplitudes of bad states decrease. This sequenceoracle followed by interferenceis the fundamental mechanism behind Grover-style amplitude amplification.

13 Amplitude Amplification and Quadratic Speedup

The quantum state can be decomposed as

$$|\Psi_{\text{init}}\rangle = \sqrt{P(e)}|\text{Good}\rangle + \sqrt{1 - P(e)}|\text{Bad}\rangle. \tag{19}$$

After $O(1/\sqrt{P(e)})$ Grover iterations, the amplitude of $|\text{Good}\rangle$ becomes dominant, yielding a quadratic speedup over classical rejection sampling.

14 Amplitude Amplification and Grover Iterations

14.1 Amplitude Parameterization

$$\theta = 2\sin^{-1}(\sqrt{P(e)}). \tag{20}$$

14.2 Grover Iteration Dynamics

$$Q^k|\psi_{\text{init}}\rangle = \cos\left(\frac{2k+1}{2}\theta\right)|\text{Bad}\rangle + \sin\left(\frac{2k+1}{2}\theta\right)|\text{Good}\rangle. \tag{21}$$

14.3 Optimal Number of Grover Iterations

$$k^* \approx \frac{\pi}{4\theta} \propto \frac{1}{\sqrt{P(e)}}. \tag{22}$$

14.4 Complexity Comparison

Method	Complexity
Classical	$O(nmP(e)^{-1})$
Quantum	$O(n2^m P(e)^{-1/2})$

15 Conclusion and Discussion

In this work, we investigated Bayesian inference under rare evidence conditions and demonstrated how quantum amplitude amplification can significantly improve sampling efficiency compared to classical rejection sampling. The study was conducted across three settings: classical simulation, noiseless quantum simulation, and execution on real quantum hardware.

15.1 Classical Rejection Sampling

In the classical setting, rejection sampling was used as the baseline inference method. As expected, the method performed poorly when the evidence probability was small. For rare evidence with probability on the order of $P(e) \approx 10^{-3}$, the vast majority of generated samples were discarded. This behavior was clearly reflected in the classical bar chart, where evidence-consistent samples appeared extremely sparsely. Obtaining even a modest number of valid samples required on the order of tens of thousands of total samples, confirming the well-known inefficiency of rejection sampling in rare-event regimes.

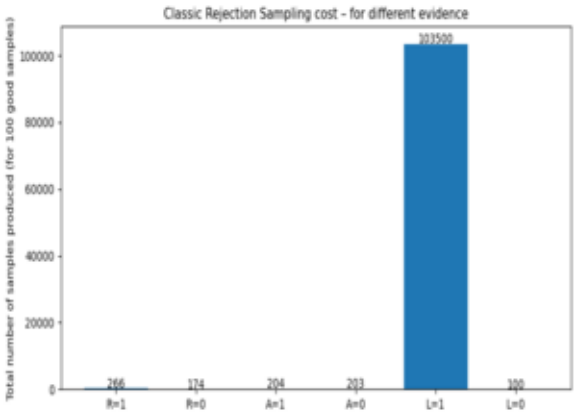


Figure 3: Default Results

This result highlights the fundamental limitation of classical inference methods: the cost scales inversely with the evidence probability $O(P(e)^{-1})$, making them impractical for rare evidence.

15.2 Quantum Simulator Results

When the same inference problem was implemented using Grover-based amplitude amplification on a noiseless quantum simulator, the results changed dramatically.
After applying the theoretically optimal number of Grover iterations, the final quantum state became almost entirely aligned with the evidence-consistent subspace. As shown in the simulator

bar chart, nearly all measured samples corresponded to valid evidence, with the success probability approaching 99%.
This behavior is expected in an ideal quantum setting:

- The simulator operates without decoherence or gate noise
- Oracle and reflection operators are applied perfectly
- Amplitude amplification performs an exact rotation toward the Good subspace

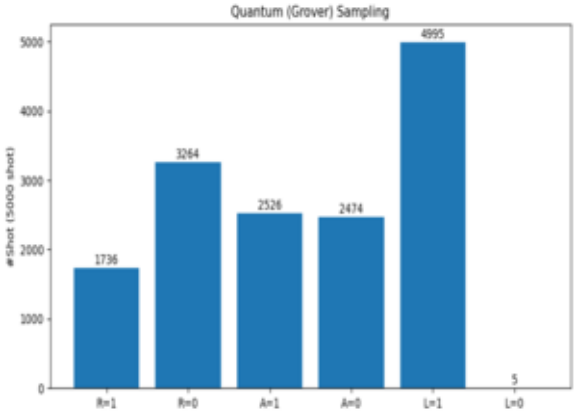


Figure 4: Simulator Results

As a result, once the optimal iteration count is reached, the simulator effectively converts a rare-event sampling problem into a near-deterministic one. This confirms the theoretical prediction of quadratic speedup and validates the correctness of the quantum circuit construction.

15.3 Real Quantum Hardware Results

When the same circuit was executed on a real IBM Quantum Processing Unit (QPU), the outcome was notably different. After the same number of Grover iterations, only approximately 27% of the measured samples satisfied the evidence condition.
While this is significantly lower than the simulators near-ideal performance, it still represents a substantial improvement over classical rejection sampling, where the success probability remained near the original evidence probability.
The reduction from the expected $\sim 99\%$ to $\sim 27\%$ (1379/5000 on graph) can be attributed to several hardware-specific factors:

- Gate noise and decoherence, which accumulate rapidly with circuit depth
- Multi-controlled gates, which are particularly error-prone on current superconducting hardware
- Over-rotation effects, where noise disrupts the precise geometric rotation required for optimal amplitude amplification
- Readout errors, which further distort the observed distribution

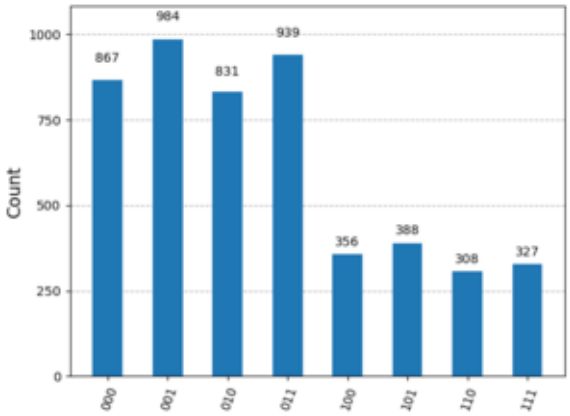


Figure 5: Hardware Results

Importantly, increasing the number of Grover iterations beyond the theoretical optimum does not solve this issue. Instead, deeper circuits exacerbate noise accumulation, causing the state to drift away from the desired subspace.

15.4 Overall Assessment

Despite hardware noise, achieving 27% evidence-consistent samples on real quantum hardware is a strong result. Compared to classical rejection sampling, this represents a meaningful practical speedup, especially in rare evidence regimes where classical methods fail entirely.
This study demonstrates that:

- Quantum amplitude amplification provides a clear theoretical advantage
- Noiseless simulators confirm the expected near-deterministic behavior
- Real QPUs already exhibit measurable benefits, even under severe noise constraints

As quantum hardware continues to improve, the gap between simulator and real-device performance is expected to narrow. The results presented here suggest that quantum Bayesian inference is not only theoretically appealing but also increasingly practical.

16 References

[1] M. de Oliveira and L. S. Barbosa. Quantum Bayesian decision-making. *Preprint*, 2023.

[2] G. H. Low, T. J. Yoder, and I. L. Chuang. Quantum inference on Bayesian networks. *Physical Review A*, 89(6):062315, 2014.

[3] P. Sakkaris. QuDot Nets: Quantum Computers and Bayesian Networks. *arXiv preprint arXiv:1601.07035*, 2016.

[4] M. Steffen, D. P. DiVincenzo, J. M. Chow, T. N. Theis, and M. B. Ketchen. Quantum computing: An IBM perspective. *IBM Journal of Research and Development*, 55(5):13:113:11, 2011.

[5] IBM Qiskit Documentation. https://qiskit.org