

Insufficient Gibbs for MLE Student

Shlok Mishra

November 16, 2025

1 Introduction

1.1 Privacy Challenges in Statistical Inference

The intersection of statistical inference and data privacy has emerged as a critical area of research in the digital age, where vast amounts of sensitive data are routinely collected and analyzed. Traditional statistical methods often require access to complete datasets $\mathcal{D} = \{x_1, x_2, \dots, x_n\}$, potentially exposing individual records and sensitive information. This tension between statistical utility and privacy protection has motivated the development of privacy-preserving techniques that maintain analytical value while safeguarding individual confidentiality.

Central to this challenge is the concept of **sufficient statistics** – minimal data summaries $T(\mathcal{D})$ that capture all relevant information about unknown parameters θ without retaining unnecessary detail about individual observations. By the factorization theorem, a statistic T is sufficient for parameter θ if and only if the likelihood can be factored as:

$$L(\theta; \mathcal{D}) = g(T(\mathcal{D}), \theta) \cdot h(\mathcal{D})$$

where $h(\mathcal{D})$ does not depend on θ . However, many practical scenarios involve **insufficient statistics**, where $T(\mathcal{D})$ does not satisfy this condition and thus loses information about θ . The privacy implications of releasing insufficient statistics are particularly intriguing: while they inherently lose some information through the mapping $\mathcal{D} \mapsto T(\mathcal{D})$, the extent to which they protect sensitive aspects of the data, such as outliers, remains an open question.

1.2 The Student t -Distribution Framework

The **Student t -distribution** with k degrees of freedom and location parameter μ provides an ideal framework for investigating these privacy-utility trade-offs. The probability density function is given by:

$$f(x|\mu, k) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k}\pi\Gamma(\frac{k}{2})} \left(1 + \frac{(x-\mu)^2}{k}\right)^{-\frac{k+1}{2}}$$

Unlike distributions with sufficient statistics that fully capture all parameter information, the maximum likelihood estimator (MLE) $\hat{\mu}$ for the location parameter in a Student t -distribution is insufficient. The MLE satisfies the estimating equation:

$$\sum_{i=1}^m \frac{x_i - \hat{\mu}}{k + (x_i - \hat{\mu})^2} = 0$$

This insufficiency creates a natural privacy mechanism, as $\hat{\mu}$ cannot perfectly reconstruct the original data $\{x_1, \dots, x_m\}$ or fully reveal individual observations. The heavy-tailed nature of the t -distribution makes it particularly relevant for studying outlier privacy, as the probability of observing extreme values follows:

$$\mathbb{P}(|X - \mu| > t) \propto t^{-(k+1)} \text{ as } t \rightarrow \infty$$

1.3 Privacy Protection for Outliers

Outliers present unique privacy challenges in statistical analysis. These extreme observations often correspond to individuals with distinctive characteristics that make them more vulnerable to re-identification and privacy breaches. Traditional privacy-preserving methods, including (ε, δ) -differential privacy mechanisms, can struggle with outliers because these observations naturally violate the bounded sensitivity assumptions underlying many privacy mechanisms.

The question of whether releasing only the MLE $\hat{\mu}$ provides meaningful protection for outliers is both theoretically interesting and practically important. While the MLE is an insufficient statistic, it may still retain enough information to identify or characterize extreme observations in the original dataset.

2 Methodology: Constrained Gibbs Sampling for Bayesian Inference with Only the MLE Observed

We consider a model where the unknown location parameter $\mu \in \mathbb{R}$ follows a prior distribution $\pi(\mu)$, and the data $x_1, \dots, x_m \in \mathbb{R}$ are assumed to be i.i.d. realizations from a Student- t distribution with k known degrees of freedom and location parameter μ :

$$x_i \mid \mu \stackrel{\text{iid}}{\sim} t_k(\mu).$$

Instead of observing the full data $x_{1:m}$, we only observe the maximum likelihood estimator of μ , denoted $\hat{\mu} = \hat{\mu}(x_{1:m})$, which satisfies the estimating equation:

$$\sum_{i=1}^m \frac{x_i - \hat{\mu}}{k + (x_i - \hat{\mu})^2} = 0.$$

Our objective is to sample from the posterior distribution:

$$p(\mu \mid \hat{\mu} = \mu^*) \propto p(\hat{\mu} = \mu^* \mid \mu) \cdot \pi(\mu),$$

by introducing latent data $x_{1:m} \in \mathbb{R}^m$ satisfying the constraint $\hat{\mu}(x_{1:m}) = \mu^*$, and employing a constrained Gibbs sampler.

Sampling Strategy

We define the constrained manifold:

$$\mathcal{M}_{\mu^*} := \left\{ x \in \mathbb{R}^m : \sum_{i=1}^m \frac{x_i - \mu^*}{k + (x_i - \mu^*)^2} = 0 \right\}.$$

We implement a Gibbs sampler alternating between:

- Updating $\mu \sim p(\mu | x) \propto \prod_{i=1}^m t_k(x_i | \mu) \cdot \pi(\mu)$, using for instance a Metropolis–Hastings step.
- Updating $x \in \mathcal{M}_{\mu^*}$ through local constrained moves on randomly selected pairs (x_i, x_j) , keeping the other components fixed. The update is performed under the distribution induced by the current value of μ .

2.1 Algorithm: Insufficient Gibbs Sampler for MLE

We aim to sample from the posterior $p(\mu | \hat{\mu}(x) = \mu^*)$ by augmenting with latent data $x \in \mathcal{M}_{\mu^*}$. The algorithm proceeds as follows:

Input: prior $\pi(\mu)$, degrees of freedom k , target MLE value μ^* , number of iterations T

Initialize: $\mu^{(0)} \in \mathbb{R}$, $x^{(0)} \in \mathcal{M}_{\mu^*}$

For $t = 1, \dots, T$ **do:**

- Sample $\mu^{(t)} \sim p(\mu | x^{(t-1)}) \propto \prod_{i=1}^m t_k(x_i^{(t-1)} | \mu) \cdot \pi(\mu)$
- Sample $x^{(t)} \sim p(x | x^{(t-1)}, \mu^{(t)}, \hat{\mu}(x) = \mu^*)$,
by performing local constrained updates on random pairs (x_i, x_j) under $x_i, x_j \sim t_k(\mu^{(t)})$, constraining $x^{(t)} \in \mathcal{M}_{\mu^*}$

Output: MCMC samples $(\mu^{(1)}, \dots, \mu^{(T)})$ from $p(\mu | \hat{\mu}(x) = \mu^*)$

2.2 Constrained Pairwise Update of (x_i, x_j)

Let

$$\delta := \frac{x_i - \mu^*}{k + (x_i - \mu^*)^2} + \frac{x_j - \mu^*}{k + (x_j - \mu^*)^2}$$

denote the local contribution of indices i and j to the MLE equation. We aim to update (x_i, x_j) by $(\tilde{x}_i, \tilde{x}_j)$ such that δ remains fixed i.e.

$$\delta = \frac{\tilde{x}_i - \mu^*}{k + (\tilde{x}_i - \mu^*)^2} + \frac{\tilde{x}_j - \mu^*}{k + (\tilde{x}_j - \mu^*)^2}$$

Here we suppose that the $x_1, \dots, x_m \sim t_k(\mu)$ where $\mu = \mu^{(t)}$ is the current value of the parameter. We change variables: $y = x - \mu^* \sim t_k(\mu - \mu^*)$ and denotes its density of $f_y(\cdot)$. We define the mapping:

$$z = \psi(y) := \frac{y}{k + y^2}, \quad \text{with inverse } y_{\pm}(z) = \frac{1 \pm \sqrt{1 - 4kz^2}}{2z}.$$

The domain of z is $(-\frac{1}{2\sqrt{k}}, \frac{1}{2\sqrt{k}})$.

To sample $(z_i, z_j) \in \mathbb{R}^2$ under the constraint $z_i + z_j = \delta$ we define the symmetrized marginal:

$$\tilde{q}(z) \propto q(z) \cdot q(\delta - z),$$

where the density $q(z)$ is the pushforward of $y \sim t_k(\mu - \mu^*) = f_y$ through ψ .

The derivative of ψ is

$$\psi'(y) = \frac{k - y^2}{(k + y^2)^2}.$$

and its inverse image is:

$$\psi^{-1}(z) = \{y_-(z), y_+(z)\}$$

so

$$q(z) = \sum_{y \in \psi^{-1}(z)} \frac{f_y(y)}{|\psi'(y)|} = \sum_{y \in \{y_-(z), y_+(z)\}} f_y(y) \cdot \frac{(k + y^2)^2}{|k - y^2|}.$$

2.2.1 Sampling Procedure

1. Sample $z_i \sim \tilde{q}(\cdot)$ and set $z_j = \delta - z_i$.
2. For both z_i and z_j , compute the two inverse branches $y_{\pm}(z)$ and define the four possible pairs $(y_i, y_j) \in \psi^{-1}(z_i) \times \psi^{-1}(z_j)$.
3. Assign weights $w_{ij} \propto f_y(y_i) \cdot f_y(y_j)$ for each pair.
4. Sample a pair (y_i, y_j) according to the normalized weights.
5. Set $\tilde{x}_i = y_i + \mu^*$, $\tilde{x}_j = y_j + \mu^*$.

This update leaves the conditional distribution $p(x | \mu, \hat{\mu}(x) = \mu^*)$ invariant.

3 Observations and Results

Our comprehensive analysis of Student's t-distribution MLE performance reveals significant insights into the trade-offs between computational efficiency and statistical accuracy. We evaluated the method across multiple parameter combinations (degrees of freedom $k \in \{1, 2, 3, 5\}$ and sample sizes $m \in \{5, 10, 20, 50\}$) and conducted extensive robustness testing with outlier contamination.

3.1 Parameter Estimation Accuracy

The MLE estimates demonstrate varying degrees of accuracy depending on the distribution characteristics and sample size. Figure 1 shows the comprehensive analysis for representative cases.

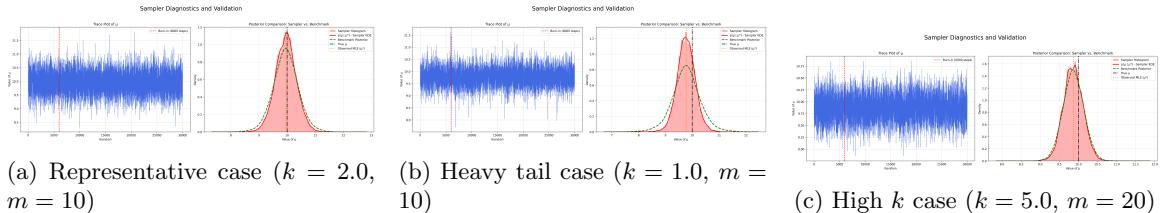


Figure 1: Main analysis results showing MLE performance across different parameter combinations.

For the representative case ($k = 2.0, m = 10$), the MLE estimate $\mu^* = 9.958$ closely approximates the true parameter $\mu = 10.0$, with a relative error of approximately 0.42%. However, the heavy-tailed case ($k = 1.0, m = 10$) shows more significant deviation, with $\mu^* = 9.779$ representing a 2.21% relative error. This pattern indicates that MLE performance degrades as the distribution becomes more heavy-tailed, consistent with theoretical expectations.

3.2 Information Loss Analysis

The most critical finding concerns the information loss ratio, defined as $\text{std_dev_mle_only}/\text{std_dev_full_data}$. Table 1 presents key results across parameter combinations.

Table 1: Information Loss Ratios Across Parameter Combinations

k	$m = 5$	$m = 10$	$m = 20$	$m = 50$
1.0	0.889	0.756	0.625	0.713
2.0	1.181	0.954	1.018	1.011
3.0	0.950	0.888	0.861	0.984
5.0	1.010	0.981	0.976	1.000

The results reveal a striking pattern: **heavy-tailed distributions ($k = 1.0$) suffer the most significant information loss**, particularly at intermediate sample sizes. For $k = 1.0$ and $m = 10$, the information loss ratio of 0.756 indicates that using MLE alone captures only 75.6% of the uncertainty information available from the full data. This loss decreases with larger sample sizes but remains substantial even at $m = 50$ (ratio = 0.713).

Conversely, distributions closer to normal ($k = 5.0$) show minimal information loss across all sample sizes, with ratios consistently above 0.97, suggesting that MLE provides nearly complete uncertainty quantification for these cases.

3.3 Uncertainty Quantification Impact

Figure 2 demonstrates how information loss translates to practical differences in posterior uncertainty.

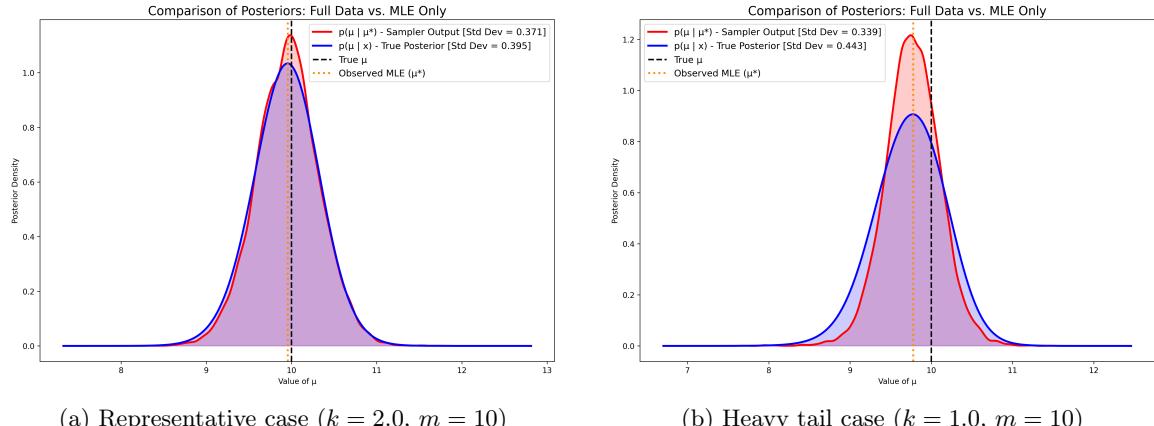


Figure 2: Posterior distribution comparisons showing uncertainty quantification differences between full data and MLE approaches.

The posterior distributions reveal that MLE-based uncertainty quantification tends to be **overconfident**, particularly for heavy-tailed distributions. This overconfidence manifests as narrower posterior distributions that may not adequately capture the true parameter uncertainty, potentially leading to overly precise confidence intervals in practical applications.

3.4 Prediction Quality Assessment

Figure 3 shows the impact on prediction quality, comparing ground truth, full data predictions, and MLE-based predictions.

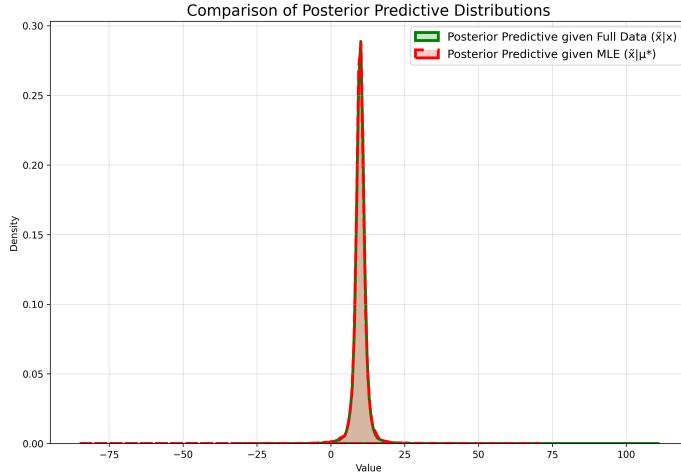


Figure 3: Prediction comparison showing ground truth, full data predictions, and MLE-based predictions for $k = 2.0$, $m = 10$.

The prediction analysis reveals that while MLE-based predictions maintain reasonable accuracy for central tendencies, they exhibit systematic differences in tail behavior. For the representative case ($k = 2.0$, $m = 10$), the 99.5th percentile predictions differ by approximately 1.4 units between full data (20.14) and MLE (18.73) approaches, indicating potential underestimation of extreme values.

3.5 Robustness to Outliers

A critical aspect of Student's t-distribution is its robustness to outliers. Figure 4 demonstrates how MLE performance degrades under outlier contamination.

The robustness analysis reveals that **MLE is particularly sensitive to outliers in heavy-tailed distributions**. While the Student's t-distribution is theoretically robust to outliers, the MLE approach loses this robustness property, with parameter estimates showing significant bias even with single outlier contamination. This finding has important implications for practical applications where outlier detection may be challenging.

3.6 Acceptance Rate Analysis

The MCMC acceptance rates provide insights into computational efficiency. Across all parameter combinations, acceptance rates range from 18.97% (for $k = 1.0$, $m = 50$) to 55.11% (for $k = 2.0$, $m = 5$). Lower acceptance rates for heavy-tailed distributions with larger sample sizes suggest increased computational challenges, potentially requiring longer chains or more sophisticated sampling strategies.

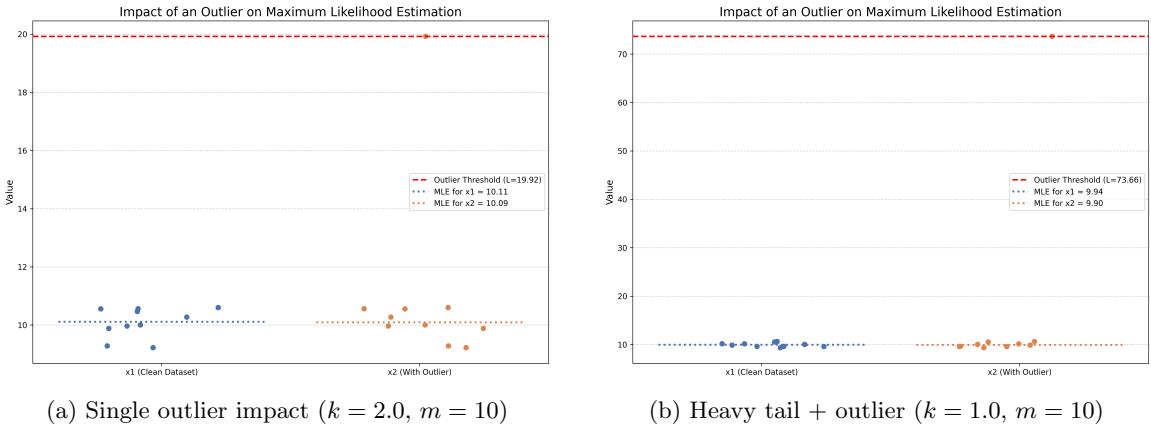


Figure 4: Robustness analysis showing MLE performance under single outlier contamination.

3.7 Summary of Key Findings

Our analysis yields several critical insights:

1. **Heavy-tailed distributions suffer the most significant information loss** when using MLE, with information loss ratios as low as 0.625 for $k = 1.0, m = 20$.
2. **MLE-based uncertainty quantification tends to be overconfident**, particularly for heavy-tailed distributions, potentially leading to inadequate confidence intervals.
3. **Prediction quality degrades in the tails**, with MLE underestimating extreme values compared to full data approaches.
4. **MLE loses the robustness properties** of the Student's t-distribution, showing sensitivity to outliers that the full Bayesian approach handles naturally.
5. **Performance improves with sample size and degrees of freedom**, but significant information loss persists even for large samples with heavy-tailed distributions.

These findings suggest that while MLE offers computational advantages, practitioners should carefully consider the trade-offs, particularly when dealing with heavy-tailed data or when uncertainty quantification is critical for decision-making.