An Instability in Variational Inference for Topic Models

Behrooz Ghorbani

Joint work with Hamid Javadi and Andrea Montanari

Stanford University
Department of Electrical Engineering

June, 2019
Problem Statement

- Statistical model:
  
  \[ X = \sqrt{\frac{\beta}{d}} WH^T + Z \]

  \[ W \in \mathbb{R}^{n \times r}, \ H \in \mathbb{R}^{d \times r} \] and \( Z \) is i.i.d Gaussian noise

- \( n, d \gg 1 \) with \( \frac{n}{d} = \delta > 0 \), where \( \delta, r \sim O(1) \)

- \( W_j \overset{i.i.d.}{\sim} \text{Dir}(\nu 1) \) and \( H_j \overset{i.i.d.}{\sim} \mathcal{N}(0, I_r) \)
Problem Statement

- **Statistical model:**

\[ X = \frac{\sqrt{\beta}}{d} WH^T + Z \]

where \( W \in \mathbb{R}^{n \times r}, H \in \mathbb{R}^{d \times r} \) and \( Z \) is i.i.d Gaussian noise.

- \( n, d \gg 1 \) with \( \frac{n}{d} = \delta > 0 \), where \( \delta, r \sim O(1) \)

- \( W_i \overset{i.i.d}{\sim} \text{Dir}(\nu 1) \) and \( H_j \overset{i.i.d}{\sim} \mathcal{N}(0, I_r) \)
Problem Statement

- Statistical model:

\[ X = \frac{\sqrt{\beta}}{d} WH^T + Z \]

where \( W \in \mathbb{R}^{n \times r}, \ H \in \mathbb{R}^{d \times r} \) and \( Z \) is i.i.d Gaussian noise

- \( n, d \gg 1 \) with \( \frac{n}{d} = \delta > 0 \), where \( \delta, r \sim O(1) \)

- \( W_i \overset{i.i.d}{\sim} \text{Dir}(\nu \mathbf{1}) \) and \( H_j \overset{i.i.d}{\sim} \mathcal{N}(0, I_r) \)
Goal: Use the posterior distribution, \( p_{H,W|X}(\cdot|X) \), to estimate \( W \) and \( H \)

**Variational Inference:** Approximate the posterior with a simpler distribution \( \hat{q} \) such that:

\[
\hat{q}(H, W) = q(H) \tilde{q}(W) = \prod_{a=1}^{d} q_a(H_a) \prod_{i=1}^{n} \tilde{q}_i(W_i)
\]

- Is the output of variational inference reliable?
(Naive Mean Field) Variational Inference

- **Goal:** Use the posterior distribution, \( p_{H,W|X}(\cdot|X) \), to estimate \( W \) and \( H \)
- **Variational Inference:** Approximate the posterior with a simpler distribution \( \hat{q} \) such that:

\[
\hat{q}(H, W) = q(H) \tilde{q}(W) = \prod_{a=1}^{d} q_a(H_a) \prod_{i=1}^{n} \tilde{q}_i(W_i)
\]

- Is the output of variational inference reliable?
Goal: Use the posterior distribution, \( p_{H,W|X}(\cdot|X) \), to estimate \( W \) and \( H \)

**Variational Inference:** Approximate the posterior with a simpler distribution \( \hat{q} \) such that:

\[
\hat{q}(H, W) = q(H) \tilde{q}(W) = \prod_{a=1}^{d} q_a(H_a) \prod_{i=1}^{n} \tilde{q}_i(W_i)
\]

Is the output of variational inference reliable?
Comparison of Two Thresholds

- $\beta_{\text{Bayes}}$: Information theoretic threshold

If $\beta < \beta_{\text{Bayes}}$, then any estimator is asymptotically uncorrelated with the truth.

If $\beta < \beta_{\text{inst}}$, $\hat{\beta} = 1 \Rightarrow$ No signal found in the data!

If $\beta > \beta_{\text{inst}}$, $\hat{\beta} \neq 1 \Rightarrow$ variational algorithm declares that it has found a signal!
Comparison of Two Thresholds

- $\beta_{\text{Bayes}}$: Information theoretic threshold
- If $\beta < \beta_{\text{Bayes}}$, then any estimator is asymptotically uncorrelated with the truth
Comparison of Two Thresholds

- $\beta_{\text{Bayes}}$: Information theoretic threshold
- If $\beta < \beta_{\text{Bayes}}$, then any estimator is asymptotically uncorrelated with the truth
- $\beta_{\text{inst}}$: Threshold for variational inference to return a non-trivial estimate
Comparison of Two Thresholds

- $\beta_{\text{Bayes}}$: Information theoretic threshold
- If $\beta < \beta_{\text{Bayes}}$, then any estimator is asymptotically uncorrelated with the truth
- $\beta_{\text{inst}}$: Threshold for variational inference to return a non-trivial estimate
- If $\beta < \beta_{\text{inst}}$, $\hat{\mathbf{w}}_i = \frac{1}{r} \mathbf{1}_r \Rightarrow$ No signal found in the data!
- If $\beta > \beta_{\text{inst}}$, $\hat{\mathbf{w}}_i \neq \frac{1}{r} \mathbf{1}_r \Rightarrow$ variational algorithm declares that it has found a signal!
Comparison of Two Thresholds

- $\beta_{\text{Bayes}}$: Information theoretic threshold
- If $\beta < \beta_{\text{Bayes}}$, then any estimator is asymptotically uncorrelated with the truth
- $\beta_{\text{inst}}$: Threshold for variational inference to return a non-trivial estimate
  - If $\beta < \beta_{\text{inst}}$, $\hat{\mathbf{W}}_i = \frac{1}{r} \mathbf{1}_r \Rightarrow$ No signal found in the data!
  - If $\beta > \beta_{\text{inst}}$, $\hat{\mathbf{W}}_i \neq \frac{1}{r} \mathbf{1}_r \Rightarrow$ variational algorithm declares that it has found a signal!
Comparison of Two Thresholds

We want $\beta_{\text{Bayes}} \approx \beta_{\text{inst}}$
Comparison of Two Thresholds

We want $\beta_{\text{Bayes}} \approx \beta_{\text{inst}}$
Comparison of Two Thresholds

We want $\beta_{\text{Bayes}} \approx \beta_{\text{inst}}$

Comparisons of $\beta_{\text{Bayes}}$ and $\beta_{\text{inst}}$
Credible intervals: Nominal coverage 90%

Empirical coverage

- $\beta = 2 < \beta_{\text{inst}}$: 0.87
- $\beta = 4.1 \in (\beta_{\text{inst}}, \beta_{\text{Bayes}})$: 0.65
- $\beta = 6 = \beta_{\text{Bayes}}$: 0.51