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:

$$\boldsymbol{X} = \frac{\sqrt{\beta}}{d}\boldsymbol{W}\boldsymbol{H}^T + \boldsymbol{Z}$$

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

- $n, d \gg 1$ with $\frac{n}{d} = \delta > 0$, where $\delta, r \sim O(1)$
- $\boldsymbol{W}_i \overset{\text{i.i.d}}{\sim} \mathrm{Dir}(\nu\mathbf{1})$ and $\boldsymbol{H}_j \overset{\text{i.i.d}}{\sim} \mathrm{N}(0, \boldsymbol{I}_r)$

Problem Statement

- Statistical model:

$$\boldsymbol{X} = \frac{\sqrt{\beta}}{d}\boldsymbol{W}\boldsymbol{H}^T + \boldsymbol{Z}$$

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

- $n, d \gg 1$ with $\frac{n}{d} = \delta > 0$, where $\delta, r \sim O(1)$
- $\boldsymbol{W}_i \overset{\text{i.i.d}}{\sim} \mathrm{Dir}(\nu\mathbf{1})$ and $\boldsymbol{H}_j \overset{\text{i.i.d}}{\sim} \mathrm{N}(0, \boldsymbol{I}_r)$

Problem Statement

- Statistical model:

$$\boldsymbol{X} = \frac{\sqrt{\beta}}{d} \boldsymbol{W} \boldsymbol{H}^T + \boldsymbol{Z}$$

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

- $n, d \gg 1$ with $\frac{n}{d} = \delta > 0$, where $\delta, r \sim O(1)$
- $\boldsymbol{W}_i \overset{\text{i.i.d}}{\sim} \mathrm{Dir}(\nu \mathbf{1})$ and $\boldsymbol{H}_j \overset{\text{i.i.d}}{\sim} \mathrm{N}(0, \boldsymbol{I}_r)$

(Naive Mean Field) Variational Inference

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

$$\hat{q}(\boldsymbol{H},\boldsymbol{W}) = q(\boldsymbol{H})\,\tilde{q}(\boldsymbol{W}) = \prod_{a=1}^{d} q_a(\boldsymbol{H}_a) \prod_{i=1}^{n} \tilde{q}_i(\boldsymbol{W}_i)$$

- Is the output of variational inference reliable?

(Naive Mean Field) Variational Inference

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

$$\hat{q}\left(\boldsymbol{H}, \boldsymbol{W}\right) = q\left(\boldsymbol{H}\right)\tilde{q}\left(\boldsymbol{W}\right) = \prod_{a=1}^{d} q_a\left(\boldsymbol{H}_a\right)\prod_{i=1}^{n}\tilde{q}_i\left(\boldsymbol{W}_i\right)$$

- Is the output of variational inference reliable?

(Naive Mean Field) Variational Inference

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

$$\hat{q}\left(\boldsymbol{H},\boldsymbol{W}\right) = q\left(\boldsymbol{H}\right)\tilde{q}\left(\boldsymbol{W}\right) = \prod_{a=1}^{d} q_a\left(\boldsymbol{H}_a\right)\prod_{i=1}^{n}\tilde{q}_i\left(\boldsymbol{W}_i\right)$$

- Is the output of variational inference reliable?

Comparison of Two Thresholds

- β_{Bayes} : Information theoretic threshold

Comparison of Two Thresholds

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

Comparison of Two Thresholds

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

Comparison of Two Thresholds

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

Comparison of Two Thresholds

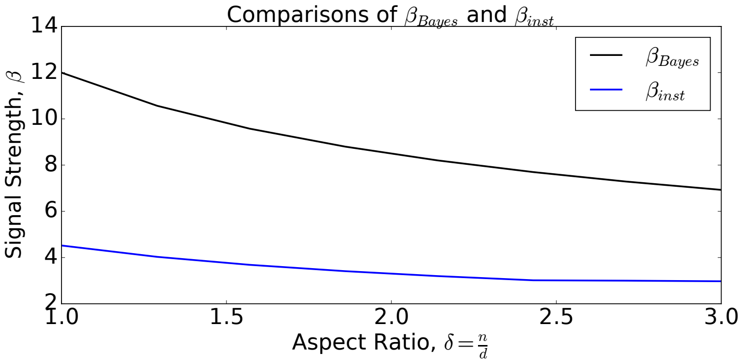
- β_{Bayes} : Information theoretic threshold
- If $\beta < \beta_{\text{Bayes}}$, then any estimator is asymptotically uncorrelated with the truth
- β_{inst} : Threshold for variational inference to return a non-trivial estimate
- If $\beta < \beta_{\text{inst}}$, $\hat{\boldsymbol{W}}_i = \frac{1}{r}\mathbf{1}_r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

Comparison of Two Thresholds

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

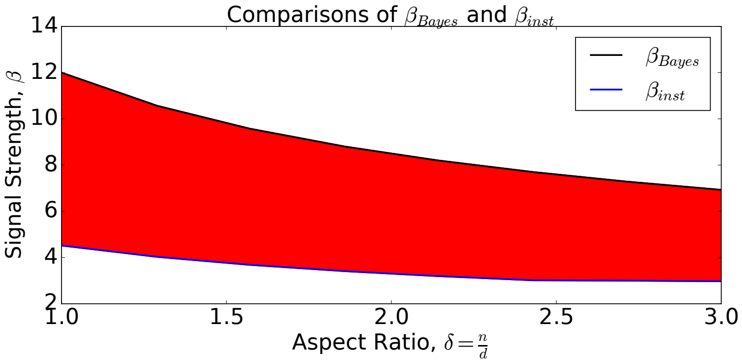
Comparison of Two Thresholds

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

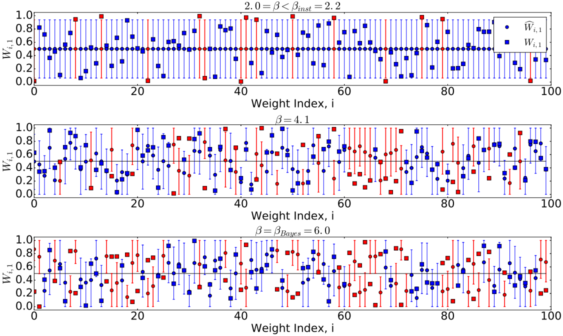


Comparison of Two Thresholds

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



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