**Abstract**

ML estimation is sensitive to outliers because it treats all data points equally. To avoid this, robust divergences were proposed.

\[ D_\beta(p(x); q(x)) = \frac{1}{\beta} \log \frac{p(x)}{q(x)} + \frac{\beta - 1}{\beta} \int q(x) \log q(x) \, dx + \frac{1}{\beta} \int f(x) \log f(x) \, dx \]

- Similarly to ML estimation, minimizing the β-divergence (or the γ-divergence) from \( p^*(x) \) to \( p(x; \theta) \) yields:
  \[ \arg \min_\theta D_\beta(p^*(x); p(x; \theta)) \]
  \[ = 0 = \frac{1}{\beta} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log p(x_i; \theta) \]

- The first term is the likelihood weighted according to the power of the probability for each data point.
- The probabilities of outliers are usually much smaller than those of inliers, and thus those weights effectively suppress the likelihood of outliers.

**Proposed method**

**Robust variational inference**

\[ \text{arg min} \int p(x; \theta) \log \frac{p(x; \theta)}{q(x)} \, dx \]

- This first term can be regarded as the expected likelihood, while the second term “regularizes” \( \theta \) to be close to the prior \( \theta^0 \).
- To enhance the robustness to outliers, we update our prior belief by Bayes' theorem and obtain the posterior:
  \[ \text{arg min} \int p(x; \theta) \log \frac{p(x; \theta)}{q(x)} \, dx \]

**Variational Bayes**

\[ \text{arg min} \int p(x; \theta) \log \frac{p(x; \theta)}{q(x)} \, dx \]

- This posterior can also be obtained by solving the following optimization problem:
  \[ \int q(x) \log q(x) \, dx = \frac{1}{\beta} \log \frac{p(x)}{q(x)} + \frac{\beta - 1}{\beta} \int q(x) \log q(x) \, dx + \frac{1}{\beta} \int f(x) \log f(x) \, dx \]

**Experimental results**

- We compare the performance of our proposed robust variational inference on UCI datasets with an existing robust variational inference method.
- We found that our method outperforms the existing methods for all datasets.
- The proposed method is capable of handling outliers more effectively.

**Main Contributions**

- To handle more complex models, we employ the optimization and variational formulation of Bayesian inference. In this formulation, the posterior model is optimized to the data under the Kullback-Leibler (KL) divergence, while it is regularized to be close to the prior.
- We propose replacing the KL divergence for data fitting to a robust divergence, such as β-divergence and γ-divergence.

**Introduction**

- Samples are generated from some unknown distribution:
  \[ \{x_i\}_{i=1}^n \sim p(x) \]
- Main body
  \[ p^*(x) = (1 - \varepsilon)p_0(x) + \varepsilon(x) \]

- In outlier-robust inference, we aim at placing an estimated probability distribution close to the main body of the unknown distribution.

**Maximum likelihood (ML) estimation**

- We estimate an unknown probability distribution \( p^*(x) \) from its independent samples \( x_{i:1}^n \).
- In ML estimation, we minimize the generalization error measured by the KL divergence from \( p^*(x) \) to a parametric model \( p(x; \theta) \).
- We approximate \( p^*(x) \) by the empirical distribution:
  \[ \hat{p}(x) = \frac{1}{n} \sum_{i=1}^n \delta(x - x_i) \]
- ML estimator is reduced to:
  \[ \arg \min_\theta D_M(p^*(x); p(x; \theta)) \]
- \( \theta \) is regarded as a random variable, following the prior \( p(\theta) \).
- We update our prior belief by Bayes' theorem and obtain the Bayesian posterior:
  \[ p(x; \theta) \sim p(x) p(\theta) \]

**Bayesian inference**

- \( \theta \) is regarded as a random variable, following the prior \( p(\theta) \).
- We update our prior belief by Bayes' theorem and obtain the Bayesian posterior:
  \[ p(x; \theta) \sim p(x) p(\theta) \]

**Main Results**

- The first term is the likelihood weighted according to the power of the probability for each data point.
- The probabilities of outliers are usually much smaller than those of inliers, and thus those weights effectively suppress the likelihood of outliers.

**Experiment on UCI datasets**

- We compare the performance of our proposed robust variational inference on UCI datasets with an existing robust variational inference method.
- We found that our method outperforms the existing methods for all datasets.
- The proposed method is capable of handling outliers more effectively.

**Main References**