**Stochastic Expectation Propagation**

Yingzhen Li¹, José Miguel Hernández-Lobato² and Richard E. Turner¹

¹University of Cambridge, ²Harvard University

{yl494, ret26}@cam.ac.uk, jmh@seas.harvard.edu

---

### From EP to Stochastic EP

**Goal:** approximate the true posterior \( q(\theta) \approx p(\theta|D) \)

\[
p(\theta|D) \propto p(D|\theta) p(\theta) \approx q(\theta) \propto p(\theta) J_1(\theta) f_1(\theta)
\]

**Idealised**

\[
p(D|\theta) \propto p(D|x_1, \theta)p(x_1|\theta)p(x_2|\theta) \approx q(\theta) \propto p(\theta)f_1(\theta) f_2(\theta) f_3(\theta)
\]

**EP**

\[
p(D|\theta) \propto p(\theta) f(\theta/N) \propto p(\theta) \approx p(\theta)
\]

**SEP**

\[
p(D|\theta) \propto p(\theta) f(\theta/N) \propto p(\theta) \approx p(\theta)
\]

**Related Algorithms**

**A) Relationships between algorithms**

- **VMP**
- **SVMP**
- **AVMP**

**B) Fixed points properties**

- **SEP**
- **DSEP**
- **AEP**

**Bayesian Neural Network**

- **Prior:** \( p_0(\lambda) = \text{Gam}(\alpha \lambda, \beta \lambda) \)
- **Likelihood:** \( p(y|x, \mathbf{w}_k) = \mathcal{N}(y, f(x, \mathbf{w}_k), \gamma_k) \)
- **Inference:** PBP* works better when under-estimating uncertainty.

**Bayesian Probit Classification**

**More Examples**

1. Clustering with MoGs
2. Odd-vs-even digit classification

**Conclusions**

- We scaled EP to large datasets with nearly identical performances
- We extended stochastic EP and related it to variational Bayes
- SEP is well suited to “big model, big data” settings

---

### An Alternative View of SEP

**Equivalent factorisation:** \( p(x|\theta) = (\prod_x p(x_n|\theta_n))^{1/N} \)

EP on the equivalent factorisation will converge to \( f(x|\theta) = f(\theta) \) that captures the average affect of likelihood terms on posterior!

**EP**

\[
p(D|\theta) \propto p(\theta) f(\theta) f(\theta/N) \propto p(\theta) \approx p(\theta)
\]

**SEP**

\[
p(D|\theta) \propto p(\theta) f(\theta) f(\theta/N) \propto p(\theta) \approx p(\theta)
\]

Work in progress: guarantees lower bound analysis of approximation error

---

### Bayesian Neural Network

- **Prior:** \( p_0(\lambda) = \text{Gam}(\alpha \lambda, \beta \lambda) \)
- **Likelihood:** \( p(y|x, \mathbf{w}_k) = \mathcal{N}(y, f(x, \mathbf{w}_k), \gamma_k) \)
- **Inference:** PBP* works better when under-estimating uncertainty.

**Bayesian Probit Classification**

- **1. Clustering with MoGs**
- **2. Odd-vs-even digit classification**

**Conclusions**

- We scaled EP to large datasets with nearly identical performances
- We extended stochastic EP and related it to variational Bayes
- SEP is well suited to “big model, big data” settings